CELLULOSE-SYNTHASE-LIKE ENZYMES AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[001] This application claims the benefit of United States Patent Application Serial Number 16/123,248, filed September 6, 2018, which is a continuation-in-part application of United States Patent Application Serial No. 14/895,059, filed December 1, 2015, which filed as a National Phase Application of PCT International Application No. PCT/IL2014/050497, international filing date June 2, 2014, which claims the benefit of United States Provisional Application Serial Number 61/831,164, filed June 5, 2013, all of which are incorporated in their entirety herein by reference; and this application also claims the benefit of Israel Patent Application Serial Number 268269, filed July 25, 2019, which is also incorporated in its entirety herein by reference.

BACKGROUND

[002] The plant kingdom produces hundreds of thousands of different small compounds, often genus or family specific. The massive structural diversity of more than 300,000 plant specialized metabolites (SMs) is generated through assorted modifications of their core structure. Plant SMs are most frequently glycosylated, and this glycosylation significantly impacts their compartmentalization, activity, solubility, stability, and toxicity. They are low molecular weight, nitrogen-containing, organic compounds, typically with a heterocyclic structure.

[003] The broad group of saponin-alkaloid compounds is widespread in plants and derived from the cytosolic mevalonic acid isoprenoid biosynthetic pathway. Saponins and steroidal alkaloids (SAs) are two large classes of SMs produced by plants.

[004] Saponins are a large group of SMs found in countless plant species and more than 100 families, representing a lipophilic triterpenoid or steroidal backbone (aglycone) decorated with one or more glycoside moieties.

[005] As suggested by their name (Latin sāpō means soap), saponins display soap like properties in aqueous solutions due to their amphipathic structure. The unique physicochemical properties of these compounds provide a broad spectrum of functions in plants including antifungal, antibacterial and insecticidal. The activity of saponins depends on the structure of the aglycone but in many cases even more on the attached sugars. For example, oleanolic and medicagenic acid (MA) derived saponins display hemolytic activity due to the presence of a carboxyl at position C-28 while Avenacin A-1 requires L-arabinose attached at the C-3 position to be fungicidal. Moreover, cholesterol-derived steroidal saponins are widespread in the plant kingdom, are highly diverse in structures, and can be either saturated (e.g., sarasapogenin) or unsaturated (e.g., diosgenin) in the C-5,6 position.

[006] Plentiful reports underline saponins' benefits to human health and medical applications

starting from anti-inflammatory and immune-boosting (adjuvant) to anti-cancer properties. UDP-glycosyltransferases (UGTs) members of the carbohydrate-active enzyme (CAZY) glycosyltransferase 1 (GT1) superfamily carry out sugar transfer reactions on saponins and all other SMs classes. The most common saccharides decorating saponins include D-glucose, D-galactose, L-arabinose, D-glucuronic acid, D-xylose, L-rhamnose and D-fucose.

[007] The presence of glucuronic acid attached at position C-3 of the sapogenin is particularly common for species of the *Caryophyllales* order, yet, the enzyme involved remains unknown.

[008] Thus, there remains an unmet need for knowledge regarding the enzymes involved in the metabolic pathway to produce saponins, including triterpenoid saponins, including knowledge of Cellulose Synthase Like Gs that attach glucuronic acid to, for example, quillaic acid. Once known, key enzymes could be regulated in a plant in order to alter the content of naturally produced products. For example, an enzyme could be over-expressed or have increased stability, thereby producing a natural sweetener at increased quantities, or an enzyme could be down regulated at the gene level to reduce production of compounds adding a bitter taste to plant products, for example quinoa. Further, there remains an unmet need for methods to produce these saponins in heterologous systems in order to provide commercial quantities of high value saponins, for example but not limited to triterpenoid saponins. These high value saponins may be used for example as, but not limited to, sweeteners, foaming agents, emulsifiers, preservatives, anti-carcinogens, hypocholesterolemic agents, anti-inflammatory agents, anti-oxidants, biological adjuvants, anti-microbial agents, insecticidal agents, antifeedants, or anti-fungal agents, or any combination thereof.

[009] In addition, some SMs, such as alkaloids, are often referred to as secondary metabolites, because they are not vital to cells that produce them, but contribute to the overall fitness of the organism. Alkaloids are low molecular weight nitrogen-containing organic compounds, typically with a heterocyclic structure, and their biosynthesis in plants is tightly controlled during development and in response to stress and pathogens.

[0010] For example, steroidal alkaloids (SAs), occasionally known as "*Solanum* alkaloids" due to their high prevalence in numerous members of the *Solanales* order, have been found to be common to numerous plants in a wide range of families. SAs have diverse structural composition and biological activity. They are low molecular weight, nitrogen-containing, cholesterol-derived organic compounds, typically with a heterocyclic structure consisting of a C-27 cholestane skeleton and a heterocyclic nitrogen component.

[0011] Estimated in the order of 1350 species, *Solanum* is one of the largest genera of flowering plants, representing about half of the species in the *Solanaceae* family (which is in the *Solanales* order). *Solanum* species include food plants, such as tomato (*Solanum lycopersicum*), potato

(*Solanum tuberosum*), bittersweet (*Solanum dulcamara*), and eggplant (*Solanum melogena*), and the *Solanaceae* also includes the *Capsicum* genus (e.g., peppers), as well as many other genera. Steroidal alkaloids are also produced by a large number of species in the *Liliaceae* family. As a result, SAs have been the subject of extensive investigations (Eich E. 2008. *Solanaceae* and *Convolvulaceae* – secondary metabolites: biosynthesis, chemotaxonomy, biological and economic significance: a handbook. Berlin: Springer).

[0012] Consisting of a C-27 cholestane skeleton and a heterocyclic nitrogen component, SAs were suggested to be synthesized in the cytosol from cholesterol. Conversion of cholesterol to the alkamine SA should require several hydroxylation, oxidation, and transamination reactions (Eich 2008, *supra*), and in most cases, further glycosylation to form steroidal glycoalkaloids (SGAs) (Arnqvist L. et al. 2003. Plant Physiol. 131:1792-1799). Glycosylation of SAs produces steroidal glycoalkaloids (SGAs), in which the added oligosaccharide moiety components directly conjugate to the hydroxyl group at C-3-beta (C-*3β*) of the alkamine steroidal skeleton (aglycone). SGA biosynthesis depends on genes encoding UDP-glycosyltransferases (UGTs) that decorate the aglycone with various oligosaccharide moieties, including D-glucose, D-galactose, L-rhamnose, D-xylose, and L-arabinose, the first two monosaccharides being the predominant units.

[0013] SGAs are produced by numerous members of the *Solanaceae* family, as well as many other families of plants. Examples of these compounds include alpha-tomatine and dehydrotomatine in tomato (*Solanum lycopersicum*), alpha-chaconine and alpha-solanine in potato (*Solanum tuberosum*), and alpha-solamargine and alpha-solasonine in eggplant (*Solanum melongena*). SGAs are also found in various types of pepper in the genus *Capsicum*. More than 100 SGAs have been identified in tomatoes (Itkin et al. , 2011, Plant Cell 23:4507-4525), and more than 50 have been identified in potatoes (Shakya and Navarre, 2008, J. Agric. Food Chem. 56:6949-6958). Eggplant also contains at least one variety of SGA (Friedmann, 2006, J. Agric. Food Chem. 54:8655-8681).

[0014] SAs and SGAs play a role in protecting plants against a broad range of pathogens and are known as phytoanticipins (antimicrobial compounds). SGAs contribute to plant resistance to a wide range of pathogens and predators, including bacteria, fungi, oomycetes, viruses, insects, and larger animals. Some SGAs in edible parts of plants are considered to be anti-nutritional compounds to humans and other mammals due to their toxic effects. For example, the SGAs alpha-chaconine and alpha-solanine are the principle toxic substances in potato, may cause gastrointestinal and neurological disorders and, at high concentrations, may be lethal to humans. Mechanisms of toxicity include disruption of membranes an dinhibition of acetylcholine esterase activity (Roddick J.G. 1989. Phytochemistry 28:2631-2634). For this reason, total SGA levels exceeding 200 mg/kg fresh weight of edible tuber are deemed unsafe for human consumption.

[0015] There is an ongoing attempt to elucidate the biosynthesis pathway of steroidal alkaloids and to control their production. U.S. Patent No. 5,959,180 discloses DNA sequences from potato which encode the enzyme solanidine UDP-glucose glucosyltransferase (SGT). Further disclosed are means and methods for inhibiting the production of SGT and thereby reduce glycoalkaloid levels in Solanaceous plants, for example potato.

[0016] Similarly, U.S. Patent Nos.7,375,259 and 7,439,419 disclose nucleic acid sequences from potato that encode the enzymes UDP-glucose:solanidine glucosyltransferase (SGT2) and β-solanine/β-chaconine rhamnosyltransferase (SGT3), respectively. Recombinant DNA molecules containing the sequences, and use thereof, in particular, use of the sequences and antisense constructs to inhibit the production of SGT2/SGT3 and thereby reduce levels of the predominant steroidal glycoalkaloids α.-chaconine and α-solanine in Solanaceous plants such as potato are also described.

[0017] Recently, three glycosyltransferases were identified that are putatively involved in the metabolism of tomato steroidal alkaloids (GLYCOALKALOID METABOLISM 1-3 (GAME1-3). More specifically, alterations in GAME1 expression modified the SA profile in tomato plants in both reproductive and vegetative parts. It is suggested that these genes are involved in the metabolism of tomatidine (the α-tomatine precursor) partially by generating the lycotetraose moiety (Itkin et al., 2011, supra).

[0018] International Patent Application Publication No. WO 00/66716 discloses a method for producing transgenic organisms or cells comprising DNA sequences which code for sterol glycosyl-transferases. The transgenic organisms include bacteria, fungi, plants and animals, which exhibit an increased production of steroid glycoside, steroid alkaloid and/or sterol glycoside compared to that of wild-type organisms or cells. The synthesized compounds are useful in the pharmaceutical and foodstuff industries as well as for protecting plants.

[0019] U.S. Patent Application Publication No. 2012/0159676 discloses a gene encoding a glycoalkaloid biosynthesis enzyme derived from a plant belonging to the family Solanaceae for example potato (*Solanum tuberosum*). A method for producing/detecting a novel organism using a gene encoding the protein is also disclosed.

[0020] U.S. Patent Application Publication No. 2013/0167271 and International Application Publication No. WO 2012/095843 relate to a key gene in the biosynthesis of steroidal saponins and steroidal alkaloids and to means and methods for altering the gene expression and the production of steroidal saponins and steroidal alkaloids.

[0021] A paper of inventors of the present invention, published after the priority date of the present invention, describes an array of 10 genes that partake in SGA biosynthesis. 5-7 of the genes were found to exist as a cluster on chromosome 7 while additional two reside adjacent in a duplicated

genomic region on chromosome twelve. Following systematic functional analysis, a novel SGA biosynthetic pathway starting from cholesterol up to the tetrasaccharide moiety linked to the tomato SGA aglycone has been proposed (Itkin M. et al., 2013 Science 341(6142):175-179).

[0022] It has also been found that the plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism (Sonawane et al. 2016. Nat. Plants 3: 16205). For example, cholesterol ((3-beta)-cholest-5-en-3-ol) is a sterol (or modified steroid), a type of lipid molecule biosynthesized by all animal cells as an essential structural component of all animal cell membranes and essential to membrane structural integrity and fluidity, enabling animal cells to function without a cell wall. It is also a precursor for the biosynthesis of steroid hormones, bile acid, and vitamin D, as well as saponins, SAs, and SGAs. Cholestanol (5-beta-cholestan-3-beta-ol; coprostanol; 5-beta-coprostanol) is a cholesterol derivative found in feces and other biological matter. Cholestanol is a 27-carbon stanol formed from the biohydrogenation of cholesterol in the gut of most higher animals (birds and mammals) by the conversion of cholesterol.

[0023] Plants make cholesterol in very small amounts, but also manufacture phytosterols (which include plant sterols and stanols, similar to cholesterol and cholestanol), which can compete with cholesterol for reabsorption in the intestinal tract, thus potentially reducing cholesterol reabsorption. Cholesterol is often used in the manufacture of medicines, cosmetics, and other applications. There is an increased interest in producing higher levels of both plant phytosterols and plant-based cholesterol.

[0024] For example, in tomato (e.g., *Solanum lycopersicum*, *Solanum pennellii*), α-tomatine and dehydrotomatine represent the major SGAs accumulating predominantly in green tissues; young and mature leaves, flower buds, skin and seeds of immature and mature green fruit. Dehydrotomatidine (i.e. tomatidenol) is the first SA aglycone formed in SGA biosynthesis which could further be hydrogenated at the C-5 position to form tomatidine. Both aglycones are further glycosylated (tetra-saccharide moiety, i.e., lycotetrose) to produce dehydrotomatine and α-tomatine, respectively. Thus, the SGA pathway branches at dehydrotomatidine for either formation of tomatidine derived SGAs or glycosylated dehydrotomatine derivatives. Notably, dehydrotomatidine and tomatidine are only different in their structures by the presence or absence of the double bond at the C-5 position. The conversion of dehydrotomatidine to tomatidine was hypothesized in the past as a single reaction catalyzed by a hypothetical hydrogenase. In most tomato plant tissues, the relative portion of dehydrotomatine as compared to α-tomatine ranges from ~2.5 - ~10%. As tomato fruit matures and reaches to the red stage, the entire pool of α-tomatine and dehydrotomatine is largely being converted to esculeosides (major SGAs) and dehydroesculeosides (minor SGAs), respectively.

[0025] In cultivated potato (e.g., *Solanum tuberosum*), α-chaconine and α-solanine are the major

SGAs sharing the same aglycone, solanidine (in which a C-5,6 double bond is present) and possess chacotriose and solatriose moieties, respectively. As there is no demissidine or demissine detected in cultivated potatoes, it was suggested that a hydrogenase enzyme able to convert solanidine to demissidine is lacking in these species. Several wild potato species (e.g. *S. demissum*, *S. chacoense*, *S. commersonii*) and their somatic hybrids (*S. brevidens* X *S. tuberosum*), predicted to contain an active hydrogenase, do produce demissidine or its glycosylated form, demissine being one of their major SGAs.

[0026] In eggplant (e.g., *Solanum melongena*), α-solamargine and α-solasonine are the most abundant SGAs derived from the solasodine aglycone (in which a C-5,6 double bond is present); while some wild *Solanum* species, e.g. *S. dulcamara* (bittersweet) produce soladulcidine or its glycosylated forms, soladulcine A and β-soladulcine (C-5,6 double bond is absent), as major SGAs from the solasodine aglycone.

[0027] In addition to SGAs, many *Solanum* species (e.g., eggplant) also produce cholesterol-derived unsaturated or saturated steroidal saponins. Unsaturated and saturated steroidal saponins are widespread in the plant kingdom, especially among monocots, e.g. the *Agavaceae* (e.g., agave and yucca), *Asparagaceae* (e.g., asparagus), *Dioscoreaceae* and *Liliaceae* families. Similar to SGAs, steroidal saponins are highly diverse in structures and could be either saturated (e.g. sarasapogenin) or unsaturated (e.g. diosgenin) in the C-5,6 position.

[0028] Cholesterol, the main sterol produced by all animals, serves as a key building block in the biosynthesis of SGAs. An array of tomato and potato **GLYCOALKALOIDMETABOLISM** (**GAME**) genes participating in core SGA biosynthesis starting from cholesterol were reported in recent years. The tomato SGAs biosynthetic pathway can be divided into two main parts. In the first, the SA aglycone is formed from cholesterol by the likely action of the GAME6, GAME8, GAME11, GAME4 and GAME12 enzymes. The second part results in the generation of SGA through the action of UDP-glycosyltransferases (UGTs): GAME1, GAME2, GAME17 and GAME18 in tomato, and STEROL ALKALOID GLYCOSYL TRANSFERASE1 (SGT1), SGT2 and SGT3 in potato.

[0029] Thus, there remains an unmet need for knowledge regarding the enzymes involved in the metabolic pathway to produce saponins, including triterpenoid saponins, including knowledge of Cellulose Synthase Like Gs that attach glucuronic acid to, for example, quillaic acid. Once known, key enzymes could be regulated in a plant in order to alter the content of naturally produced products. For example, an enzyme could be over-expressed or have increased stability, thereby producing a natural sweetener at increased quantities, or an enzyme could be down regulated at the gene level to reduce production of compounds adding a bitter taste to plant products, for example quinoa. Further, there remains an unmet need for methods to produce these saponins in heterologous systems in order

to provide commercial quantities of high value saponins, for example but not limited to triterpenoid saponins. These high value saponins may be used for example as, but not limited to, sweeteners, foaming agents, emulsifiers, preservatives, anti-carcinogens, hypocholesterolemic agents, anti-inflammatory agents, anti-oxidants, biological adjuvants, anti-microbial agents, insecticidal agents, antifeedants, or anti-fungal agents, or any combination thereof.

[0030] In addition, the demand for higher food quantities and food with improved quality continues to increase. Improved nutritional qualities as well as removal of antinutritional traits are both of high demand. In the course of crop domestication, levels of anti-nutrients were reduced by breeding, However, many crop plants still contain significant amount of antinutritional substances, particularly steroidal glycoalkaloids.

[0031] Alternatively, the ability to manipulate the synthesis of these SGAs would provide the means to develop, through classical breeding or genetic engineering, crops with modified levels and composition of SGAs, conferring on the plant an endogenous chemical barrier against a broad range of insects and other pathogens.

[0032] In addition, there is a demand both for plant-based cholesterols and, conversely, for plants with increased levels of phytocholesterols or other phytosterols.

[0033] Thus, there is also a demand for, and it would be highly advantageous to have, means and methods for controlling the production of saponins, steroidal alkaloids, and steroidal glycoalkaloids with beneficial, particularly therapeutic, effects.

[0034] The disclosure provided herein below, meets these unmet needs by describing in some embodiments, at least the knowledge that plant SMs, more specifically saponins, are not merely glycosylated by UGT 1 family enzymes. It appears that proteins related to Cellulose Synthases (CESA), renowned for their role in primary cell wall biosynthesis, enable the attachment of glucuronic acid at the C-3 position of saponins. Metabolic evolution through neofunctionalization to a Cellulose Synthase Like (CSL) enzyme possessing a new form of triterpenoid glycosyl transferase activity was accompanied by altering its subcellular localization, directing it to the endoplasmic reticulum (ER). Modulation of this protein also resulted in strict control of the production of steroidal saponins and steroidal glycoalkaloids, while gene silencing resulted in accumulation of cholesterol pools.

[0035] The disclosure provided herein below, also meets these unmet needs by describing in some embodiments, at least the knowledge of key genes and enzymes in the biosynthesis pathway converting cholesterol to steroidal saponins, triterpenoid saponins, steroidal alkaloids, and steroidal glycoalkaloids.

SUMMARY

[0036] According to one aspect, provided herein is a genetically modified cell having increased expression of at least one heterologous gene compared to a corresponding unmodified cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said genetically modified cell comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell.

[0037] According to another aspect, provided herein is a genetically modified plant comprising at least one cell having altered expression of at least a cellulose synthase like G (CSLG) gene compared to the expression of CSLG in a corresponding unmodified plant, and wherein the genetically modified plant has an altered content of at least one steroidal alkaloid, a derivative thereof a metabolite thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[0038] According to still another aspect, provided herein is a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin in a genetically modified cell, the method comprising: (a) introducing an at least one heterologous gene into said cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said heterologous gene is optionally comprised in a vector; and (b) expressing said at least one heterologous gene in said cell; wherein said cell comprises an increased content of at least one steroidal alkaloid, at least one steroidal saponin, or at least one triterpenoid saponin compared to a corresponding unmodified cell.

[0039] According to yet another aspect, provided herein is a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or a plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising: (a) transforming said at least one plant cell with at least one silencing molecule targeted to a nucleic acid gene sequence encoding a Cellulose Synthase Like G (CSLG) enzyme; or (b) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any

combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; wherein expression of the gene encoding the CSLG enzyme is reduced in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell, wherein the plant comprising said genetically modified cell comprises reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[0040] According to yet another aspect, provided herein is a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising: (a) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; and (b) expressing said nucleic acid encoding said CSLG; wherein the plant comprising said genetically modified cell comprises increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

BRIEF DESCRIPTION OF THE DRAWINGS

[0041] The subject matter related to methods of production of steroidal glycoalkaloids, steroidal saponins, and triterpenoid saponins and uses thereof is particularly pointed out and distinctly claimed in the concluding portion of the specification. The methods of production of these steroidal glycoalkaloids, steroidal saponins, and triterpenoid saponins and uses thereof, however, both as to organization and method of operation, together with objects, features, and advantages thereof, may best be understood by reference to the following detailed description when read with the accompanying drawings in which:

[001] **FIGURE 1** shows the proposed biosynthetic pathway of steroidal glycoalkaloids in the

triterpenoid biosynthetic pathway in Solanaceous plant from cholesterol toward α-tomatine. Dashed and solid arrows represent multiple or single enzymatic reactions in the pathway, respectively.

[002] **FIGURE 2** summarizes the coexpression analysis of steroidal alkaloid-associated genes in Solanaceous plants. Shared homologs of coexpressed genes for 'baits' from tomato (*SlGAME1* and *SlGAME4*) and potato (*StSGT1* and *StGAME4*). Continuous (r-value > 0.8) and dashed (r-value > 0.63) lines connect coexpressed genes. *, located in the tomato or potato chromosome 7 cluster. St, *Solanum tuberosum*; Sl, *S. lycopersicum*. Background of gene names corresponds to bait they were found to be coexpressed with (legend above). SP, serine proteinase; PI, proteinase inhibitor; UPL, ubiquitin protein ligase; ELP, extensin-like protein; PK, protein kinase; SR, sterol reductase; RL, receptor-like.

[003] **FIGURE 3** presents schematic map of genes identified in the duplicated genomic regions in tomato and potato and their coexpression. Coexpression with *GAME1/SGT1* (chromosome 7) and *GAME4* (chromosome 12) as baits in either potato or tomato are presented in a form of a heatmap (**Tables 5-8**). Specific gene families are indicated by dark arrows while members of other gene families are in white arrows.

[004] **FIGURES 4A-4H** shows functional analysis of tomato GAME genes. (**FIGURE 4A**) *GAME8*-silenced transgenic (RNAi) leaves accumulated 22-(R)-hydroxycholesterol compared to wild type. (**FIGURE 4B**) An array of cholestanol-type steroidal saponins (STSs) accumulates in *GAME11* VIGS-silenced leaves. (**FIGURE 4C**) An STS (m/z = 753.4) accumulates in *GAME12* VIGS- leaves. (**FIGURE 4D**) Tomatidine, the steroidal alkaloid aglycone, accumulates in *GAME1*-silenced transgenic leaves. (**FIGURES 4E to 4H**) Enzyme activity assays of the 4 recombinant tomato GAME glycosyltransferases.

[005] **FIGURES 5A-5D** show solanine/chaconine levels in peels of tuber of potato plant lines with altered expression of *GAME9* compared to wild type plants. Solanine (**FIGURE 5A**) and chaconine (**FIGURE 5B**) level in tubers of *GAME9* silenced plant; Solanine (**FIGURE 5C**) and chaconine (**FIGURE 5D**) levels in tubers of *GAME9* overexpressing plants.

[006] **FIGURE 6** shows solanine/chaconine levels in leaves of potato plant lines with either silenced (RNAi) or overexpressed (OX) *GAME9* compared to wild type plants.

[007] **FIGURE 7** shows tomatine levels in leaves of tomato plant lines with either silenced (RNAi, line 5871) or overexpressed (OX, line 5879) *GAME9* compared to wild type plants.

[008] **FIGURES 8A-8D** show the effect of silencing of GAME11 dioxygenase in tomato. (**FIGURE 8A**) α-tomatine levels in leaves (m/z = 1034.5) (**FIGURE 8B**) cholestanol-type steroidal saponins (STS) in leaves (m/z = 1331.6, 1333.6, 1199.6, 1201.6 (major saponins)).

(**FIGURE 8C**) MS/MS spectrum of m/z = 1331.6 (at 19.28 min.). (**FIGURE 8D**) The fragmentation patterns of the saponin eluted at 19.28 min. and accumulating in *GAME11*-silenced leaves. Corresponding mass signals are marked with an asterisk on the MS/MS chromatogram in **FIGURE 8C**.

[009] **FIGURES 9A-9E** show metabolites extracted from *GAME18*-silenced mature green tomato fruit. Peaks of newly accumulating compounds corresponding to the γ-tomatine standard (m/z = 740.5) (**FIGURES 9A-C**), and γ-tomatine pentoside (m/z = 872.5) (**FIGURES 9D-E**) are shown.

[010] **FIGURES 10A-10D** show the effect of silencing of *GAME12* transaminase in tomato. (**FIGURE 10A**) accumulation of a furastanol-type STS. (**FIGURES 10B-10C**) *GAME12*-silenced leaves accumulate an STS (m/z=753.4), while it exists in only minor quantities in WT leaf. (**FIGURE 10D**) MS/MS spectrum of m/z=753.4 at 19.71 min. with interpretation of the fragments.

[011] **FIGURES 11A-11D** show the effect silencing of *GAME8* in tomato plants. *GAME8*-silenced leaves accumulated 22 -(S) and -(R)-cholesterol (**FIGURE 11A**). Chromatograms (mass range 172.5-173.5) acquired via EI-GC/MS, MS spectra and structures (tri-methyl-silyl derivatives) of the compounds are shown. Commercial standards of 22-(R)- (**FIGURE 11B**) and 22-(S)-cholesterol (**FIGURE 11C**) were used to verify the putative identification. (**FIGURE 11D**) *GAME8*-silenced line accumulates both isomers in comparison to WT (Q).

[012] **FIGURE 12** shows the phylogenetic tree of GAME genes in the plant CYP450 protein family. The numbers on the branches indicate the fraction of bootstrap iterations supporting each node.

[013] **FIGURE 13** shows a proposed expanded biosynthetic pathway in *Solanaceous* plants from Cycloartenol (Part I), through Cholesterol (Part II), through Tomatidine (Part III), through Steroidal Glycoalkaloids including α-tomatine to Lycoperosides/Esculeoside (Part IV). Dashed arrows represent multiple enzymatic reactions in the pathway.

[014] **FIGURES 14A-14D** show an overview of SGA biosynthesis in (**A**) tomato, (**B**) potato, and (**C**) eggplant. **Figure 14D** shows a proposed portion of an SGA biosynthetic pathway in *Solanaceous* plants, from tomatidine to the production of tomatidine 3-O-glucuronide, wherein a GAME15 cellulose synthase like G enzyme catalyzes the addition of a glucuronic acid at the position of the hydroxyl group of the tomatidine.

[015] **FIGURES 15A-15C** show major SGA levels in (**A**) leaves and (**B**) green fruit and (**C**) red fruit of wild type (non-transformed) and *GAME15*-RNAi tomato lines determined by LC-MS. #21, #22 and #23 are three independent *GAME15*-RNAi transgenic tomato lines. Values indicate means of three biological replicates ± standard error. Asterisks indicate significant changes from wild-

type samples as calculated by a Student's *t*-test (*P-value < 0.05; **P-value < 0.01; ***P-value < 0.001).

[016] **FIGURE 16** shows levels of α-solanine and α-chaconine in leaves of *GAME15*-RNAi lines as determined by LC-MS. #1, #2 and #3 are three independent *GAME15i* transgenic potato lines. Values represent mean ± standard error (n=3). Student's *t*-test was used to assess whether the transgenic lines significantly differ from wild-type plants: (*P-value < 0.05; **P-value < 0.01; ***P-value < 0.001).

[017] **FIGURE 17** shows the cholesterol content of tomato leaves derived from *GAME15* silenced plants. Values represent mean of three biological replicates ± standard error. Asterisks indicate significant changes in leaves of the three independent transgenes (#21, #22 and #23) as compared to wild-type leaves (i.e. non-transformed) calculated by a Student's *t*-test (*P-value < 0.05; **P-value < 0.01; ***P-value < 0.001). Epicholesterol was used as an internal standard in sample preparations and relative cholesterol level is expressed as ratios of cholesterol peak areas in sample compared to internal standard. The analysis was performed using GC-MS.

[0042] **FIGURE 18** are liquid chromatography-electrospray ionization-quadrupole-time of flight-mass spectrometry (LC-ESI-QTOF-MS) scans of spinach saponins. The upper scan shows the total ion chromatogram (TIC) of spinach leaf extract acquired in negative ion mode. Chromatogram fragment displayed in red shows saponins present in the sample. The lower scan presents an expanded region of the chromatogram with main saponins being numbered. The numbers represent different Yossosides as listed in **Table 13** presented in **Example 15**.

[0043] **FIGURES 19A-19P** provide Tandem mass spectrometry (MS/MS) results and structures of Yossoside IV (**FIGURES 19A-19D**); of Yossoside Va (**FIGURES 19E-19G**); of Yossoside V (**FIGURES 19H-19J**); of Yossoside IX (**FIGURES 19K-19M**); and of Yossoside X (**FIGURES 19N-19P**). Details of the MS/MS figures showing results and structures are as follows: (**FIGURE 19A**) Mass fragments originating from [M-H]11963.56 = - *m/z* range for 300-1300; (**FIGURE 19B**) Mass fragments originating from the cleavage of glucuronic acid residue - *m/z* range for 50-1900; (**FIGURE 19C**) Structure of Yossoside IV, arrows indicate fragmentation patterns; (**FIGURE 19D**) Table with assignments of MS/MS peaks originating from the fragmentation of glucuronic acid bound to MA; (**FIGURE 19E**) Mass fragments originating from [M-H]1305.57190 = - *m/z* range for 300-1400; (**FIGURE 19F**) Mass fragments originating from the cleavage of glucuronic acid residue - *m/z* range for 50-1900; (**FIGURE 19G**) Structure of Yossoside Va, arrows indicate fragmentation patterns. Position of acetyl group (highlighted in red) is putative.; (**FIGURE 19H**) Mass fragments originating from [M-H]1305.5737 = - *m/z* range for 300-1400; (**FIGURE 19I**) Mass fragments originating from the cleavage of glucuronic acid residue - *m/z* range for 50-1900; (**FIGURE 19J**)

Structure of Yossoside V, arrows indicate fragmentation patterns; (**FIGURE 19K**) Mass fragments originating from [M-H]1407.6060 = - *m/z* range for 300-1500; (**FIGURE 19L**) Mass fragments originating from the cleavage of glucuronic acid residue - *m/z* range for 50-1900; (**FIGURE 19M**) Structure of Yossoside IX, arrows indicate fragmentation patterns. Structures and places of attachment of pentoses are putative.; (**FIGURE 19N**) Mass fragments originating from [M-H]1437.6150 = - *m/z* range for 300-1500; (**FIGURE 19O**) Mass fragments originating from the cleavage of glucuronic acid residue - *m/z* range for 50-1900; and (**FIGURE 19P**) Structure of Yossoside X, arrows indicate fragmentation patterns. Structure and attachment of pentose is putative.

[0044] **FIGURES 20A-20C** show that the biosynthetic pathway of spinach saponins comprising co-expressed saponin β-amyrin synthase (SOAP) genes. (**FIGURE 20A**) The complete biosynthetic pathway leading to the production of Yossoside V; an acetylated triterpenoid saponin containing glucuronic acid attached at C-20 position (highlighted in red). Each product of SOAP enzymes activity is highlighted in green. SOAP enzymes are those enzymes involved in saponin production in spinach. (**FIGURE 20B**) Gene co-expression network. Dots in red represent baits (SOAP1, SOAP2 and SoCYP716A268v2), circle with green dots depicts genes co-expressed with all the baits. Enlarged and annotated dots represent genes silenced by Virus Induced Gene Silencing (VIGS) in spinach. (**FIGURE 20C**) Extracted ion chromatograms (EIC) of oleanolic acid, augustic acid and medicagenic acid [m/z = 455.205 (OA); m/z = 471.205 (AA); m/z = 501.202 (MA); in negative ion mode] from plants transiently expressing SOAP1-4, compared to plants expressing SOAP1-20, SOAP1-2 alone and to control (plants infiltrated with Agrobacterium rhizogenes harboring empty vector - EV).

[0045] **FIGURE 21** presents important Heteronuclear Multiple Bond (HMBC) correlations of the hydrogens of Yossoside V. Black arrows represent HMBC correlations of the six methyl groups of Yossoside V. Red arrows represent HMBC correlations of other important hydrogens. Numbers correspond to carbon atoms in medicagenic acid and sugar moieties. For more details see **Table 1**. **Table 1** presents spectral data summarized in tables that are split into two parts. The first part corresponds to atoms of medicagenic acid (aglycone part) while the second part corresponds to atoms of sugar moieties.The structures of Yossoside V and medicagenic acid 3-O-glucuronide with numbered atoms can be found in **FIGURE 21** and **FIGURE 35D**, respectively. In **Table 1**: ax (axial) and eq (equatorial) depict protons in ring structures of aglycone and sugars.

[0046] **Table 1: 1H and 13C NMR spectral data for Yossoside V (3-O-[β-D-xylopyranosyl-(1–>3)-β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-4-acetyl-β-D-fucopyranosyl]-medicagenic acid) and for Medicagenic acid-3-O-glucuronide.**

PCT/IL2019/051000

Yossoside V					
aglycone part			sugars		
Atoms No.	^{13}C shift (ppm)	^{1}H shift (ppm)	Atoms No.	^{13}C shift (ppm)	^{1}H shift (ppm)
1	44.77	2.08 (eq)	β-D-Fuc		
		1.24 (ax)	1	94.94	5.38 (ax)
2	70.68	4.30	2	74.58	3.74 (ax)
3	86.47	4.05 (ax)	3	74.57	3.88 (ax)
4	53.01		4	75.13	5.06 (eq)
5	53.01	1.57 (ax)	5	71.06	3.83 (ax)
6	21.53	1.61	6	16.37	1.06
		1.16	acetyl		
7	33.65	1.35 (eq)	CO	172.69	
		1.49 (ax)	Methyl	20.78	2.14
8	40.96				
9	49.44	1.57 (ax)	α-L-Rha		
10	37.24		1	101.48	5.33 (eq)
11	24.68	1.93 (eq)	2	71.66	3.90 (eq)
		1.99 (ax)	3	72.04	3.83 (ax)
12	123.51	5.26	4	84.52	3.53 (ax)
13	144.57		5	68.74	3.80 (ax)
14	43.09		6	18.28	1.32
15	28.96	1.18 (eq)			
		1.58 (ax)	β-D-Glc		
16	23.92	1.60 (eq)	1	106.4	4.47 (ax)
		2.04 (ax)	2	76.14	3.18 (ax)
17	48		3	78.28	3.34 (ax)
18	42.8	2.81	4	71.46	3.27 (ax)
19	47.22	1.13	5	78.09	3.25 (ax)
		1.71	6a	62.85	3.66
20	31.4		6b	62.85	3.84
21	34.77	1.22 (eq)			
		1.38 (ax)	β-D-GlcA		
22	33.01	1.56 (eq)	1	104.79	4.39
		1.75 (ax)	2	68.93	3.81
23	182.07		3	70.94	3.81
24	13.57	1.37 (ax)	4	74.31	3.43
25	17.08	1.24 (ax)	5	nd	nd
26	17.66	0.78 (ax)	6	nd	
27	26.24	1.14 (ax)			
28	177.9		β-D-Xyl		
29	33.45	0.89 (eq)	1	105.38	4.55 (ax)
30	24.1	0.92 (ax)	2	75.17	3.25 (ax)
			3	77.31	3.31
			4	70.89	3.48
			5a	66.85	3.87
			5b	66.85	3.2

PCT/IL2019/051000

Medicagenic acid 3-*O*-glucuronide		
aglycone part		
Atoms No.	^{13}C shift (ppm)	^{1}H shift (ppm)
1	43.72	1.24 (ax)
		2.09 (eq)
2	69	4.32
3	87.14	4.07 (ax)
4	52.97	
5	51.73	1.62 (ax)
6	20.49	1.29
		1.62
7	32.44	1.24 (eq)
		1.56 (ax)
8	39.44	
9	48.18	1.60 (ax)
10	36.12	
11	23.31	1.89 (eq)
		2.02 (ax)
12	121.05	5.22
13	145.36	
14	41.9	
15	27.84	0.95 (eq)
		1.87 (ax)
16	23.2	1.57 (eq)
		1.88 (ax)
17	nd	
18	42.07	2.90
19	46.73	1.08
		1.65 (eq)
20	30.43	
21	34.12	1.12 (eq)
		1.34 (ax)
22	32.94	1.49 (eq)
		1.72 (ax)
23	184.8	
24	13.46	1.36 (ax)
25	15.8	1.26 (ax)
26	16.97	0.88 (ax)
27	25.15	1.14 (ax)
28	nd	
29	32.56	0.87 (eq)
30	22.81	0.95 (ax)

glucuronic acid		
Atoms No.	^{13}C shift (ppm)	^{1}H shift (ppm)
β-D-GlcA		
1	103.14	4.41 (ax)
2	74.02	3.27 (ax)
3	75.95	3.40 (ax)
4	74.52	3.60 (ax)
5	72.09	3.39 (ax)
6	175.6	

[0047] **FIGURES 22A and 22B** show that the genes involved in first steps of aglycone biosynthesis are located in close vicinity on the chromosome. (**FIGURE 22A**) Schematic representation of the chromosomal region with genes involved in oleanolic acid biosynthesis. Semitransparent gray boxes contain SOAP1 (bAS) (SEQ ID NO: 422), CYP716A268 (SOAP2; SEQ ID NO: 46) and its truncated duplication CYP716A268v2 (SEQ ID NO: 47). Genes marked in dark grey are not directly involved in the biosynthesis of triterpenoid saponins. (**FIGURE 22B**) Expression pattern of SOAP1, SOAP2 and CYP716A268v2 among five spinach transcriptomes; SO4WOLLD - mature leaf from four week old plant grown in long day; SO8WOLSL - mature leaf from eight week old plant grown in short day; SO4WYLLD - young leaf from four week old plant grown in long day; SO8WOLSL - young leaf from eight week old plant grown in short day; SO4WFBLD - flower bud from four week old plant grown in long day. For details see Materials and Methods.

[0048] **FIGURES 23A-23D** present data providing functional characterization of SOAP1 (bAS) and SOAP2 (CYP7123A2238) by silencing (VIGS) in spinach. (**FIGURE 23A**) Chlorotic phenotype observed in plant with silenced magnesium-chelatase subunit H (CHLH). CHLH was used as a control and a marker of VIGS gene silencing. In general, this gene is not directly related to SOAP genes and saponin biosynthesis. EV stands for empty vector and represents a control leaf. (**FIGURE 23B**) Real-time qPCR analysis of SOAP1 and SOAP2 expression in silenced plants. Values (±SD) represent mean of three independent biological experiments. Statistically significant differences compared with control plants (CHLH silenced) are indicated; ***$P < 0.001$. (**FIGURE 23C**) Aligned gas chromatography–mass spectrometry (GC-MS) extracted ion chromatograms (EICs) of P-amyrin [m/z = 218, main fragment] from the plants with silenced SOAP1 or SOAP2 compared to control (CHLH). P-Amyrin can be detected only in plants with silenced SOAP2 where its further conversion is blocked. (**FIGURE 23D**) Aligned liquid chromatography–mass spectrometry (LC-MS) EICs of Yossoside IV [m/z = 12233.23, (-)-mode] from plants with silenced SOAP1 or SOAP2 compared to control (CHLH). Content of Yossoside IV was reduced to 40 and 30% in SOAP1 and SOAP2 silenced lines, respectively.

[0049] **FIGURES 24A and 24B** present data providing functional characterization of SOAP1 (bAS; SEQ ID NO: 45) and SOAP2 (CYP2416A268; SEQ ID NO: 46) by heterologous expression in *N. benthamiana*. (**FIGURE 24A**) Aligned GC-MS extracted ion chromatograms (EICs) of p-amyrin [m/z = 218, main fragment] from the transient expression of SOAP1 compared to empty vector (EV) alone and an authentic standard. Only chromatograms from *in planta* experiments are to scale. (**FIGURE 24B**) Aligned GC-MS EICs of oleanolic acid [m/z = 203, main fragment] from the transient expression of SOAP1 and SOAP2 compared to empty vector (EV) alone and an authentic standard. Only chromatograms from *in planta* experiments are to scale. For details of GC-MS

analysis see Materials and Methods.

[0050] **FIGURE 25** presents data showing liquid chromatography-mass spectroscopy (LC-MS) analysis of selected saponins in spinach with (VIGS) silenced cytochrome P450 candidate genes. Relative content (normalized peak area) of selected saponin compounds in plants with silenced SOAP3 (CYP72A655; SEQ ID NO: 51), SOAP4 (CYP72A654; SEQ ID NO: 53), and additional three other CYPs (p450 cytochrome genes); for accession numbers see Table 11. Values (±SD) represent mean of three independent biological experiments. Statistically significant differences compared with control plants (CHLH silenced) are indicated; *P < 0.05, **P < 0.01.

[0051] **FIGURES 26A-26L** present characterization of saponins accumulating in spinach following silencing of cytochrome P450 genes (VIGS). (**FIGURES 26A, 26D, 26G, and 26J**) Relative content (normalized peak area) of selected triterpenoid saponin compounds in spinach leaves with silenced SOAP3 (CYP72A655; SEQ ID NO: 51), SOAP4 (CYP72A654; SEQ ID NO: 53) and additional three cytochrome p450 genes; for accession numbers see Table11. Values (±SD) represent mean of three independent biological experiments. Statistically significant differences compared with control plants (CHLH silenced) are indicated; *P < 0.05, **P < 0.01. AA- augustic acid (**FIGURES 26B, 26E, 26H, and 26K**) Tandem mass spectroscopy (MS/MS) spectra with marked main fragments. (**FIGURES 26C, 26F, 26I, and 26L**) Putative structures of saponin compounds based on accurate mass measurement, fragmentation patterns, and literature. For more details see Table 16.

[0052] **FIGURE 27** presents GC-MS analysis of saponins' aglycones in spinach plants with silenced CYPs P450. EICs of triterpenoid aglycones [m/z = 203 and 262; common fragments for all amyrin-type aglycones] released from the saponins accumulated in plants with silenced SOAP3 alone, SOAP4 alone, and CYP2 compared to control (CHLH silenced) and authentic standards of oleanolic and medicagenic acid. Scale of each EIC was normalized to the most abundant signal. Numbers from 1-7 represent names of detected compounds, 1 - oleanolic acid; 2 - hederagenin; 3 - augustic acid; 4 - gypsogenin; 5 - bayogenin; 6 - unknown; 7 - medicagenic acid, respectively.

[0053] **FIGURES 28A-28C** presents the analysis and putative characterization of aglycones in spinach with silenced CYPs P450. (**FIGURE 28A**) Relative content of detected sapogenins in plants with silenced SOAP3, SOAP4, CYP2 and CHLH as control. (**FIGURE 28B**) Structures of characterized aglycones. Numbers from 1-7 represent names of detected compounds, 1 - oleanolic acid; 2 - hederagenin; 3 - augustic acid; 4 - gypsogenin; 5 - bayogenin; 6 - unknown; 7 - medicagenic acid, respectively. (**FIGURE 28C**) Fragmentation patterns of described compound and comparison with fragmentation patterns of authentic standards of oleanolic acid and medicagenic acid.

[0054] **FIGURE 29** presents LC-MS analysis of selected saponins in spinach with silenced (VIGS) for glycosyltransferase genes. Relative content (normalized peak area) of selected saponin

compounds in plants with silenced SOAP6-9 (SEQ ID NOs: 55, 57, 59, and 61, respectively), *SoGT1* (SEQ ID NO: 117 [gene] and SEQ ID NO: 118 [polypeptide]) and *SoGT2* (SEQ ID NO: 119 [gene] and SEQ ID NO: 120 [polypeptide]); for accession numbers see Table 11. Sequences used for the glycosyl transferases are provided below.

[0055] Spinach GT1 nucleic acid gene sequence is set forth as follows: atggccgcctcaaacaaagagcaaagcaagctacatattgctatgtttccttggtttgcatatggtcatataaacccattcatccagctctctaacaa gctttcctcccatggcatccaaatctctttcttctcaataccaggaaacattgatcgtatcaaatcctcccttaatctctcacctcccaaccagctcatc cccctcactattcccccaactgaaggactctctcccaactttgacagcagctctgaagtaacacctcaaactgctcagcttctcacactagcacttg atcaaatgcagccccaagtcaaagctctattccctcacccccagccacaagttatcctctttgattttgcatatcactggcttccctcagtagcttctg aactaggcatcaaagctgttcatttcaatacattcccagctgttatcaattcatatctcactgtcccttcaagaatgactgatccaaataaaccaccga catttgaggacttgaagaaccctcctcaaggctatcccaaaacctcaaccgcctcagtgaaaaccttcgaagctcaagattacctattccttttcaa gagtttcgatggcggaccgtgccatttcgaaaagatattggcattcacaaacagctgtgatgctatactttacaggacctgcaatgaaatagaagg tccattcatagattacttcaagacccaaataaataaaccactgcttttagctggcccaagtgttcctctaccaccctctggtgaactggatgaaaaat gggagatgtggttaggtaaatttcctgaaaagtcagtcatatactgcagcttcggaagcgagacatacttgaatgatgctcagattcaggagctta cacttgggttggagctcactggtctgccctttatcttggttttgaattttggaacaagtaacagcaccgatgcccacaataagctagaagcatcatta ccagaaggatttagagagagaatcaaagacaggggcgttctgcatacaggatgggtgcaacagcaaaacattttagcacacagaagcatagg atgctttcttactcacgcagggttcagctctgtaatagagggtattgtgaatgactgtcaattagcatttctacctctaaaggctgaccagtttatgatc gctaagctatttagtggggatctgaaagcaggggtggaggtaaatcgaagagatgaagatgggtcttttgccaaagaagatattttcgaagcaat aaagacgatcatggtggatactgataaagaaccaagtagatccatcagggagaatcatagcaattggagaaagtttttgatgaataaggagattg aagctagttatattgcaaatttagctcatgaactcaaggcattggttcaaaaagcttag (SEQ ID NO: 117). Spinach GT1 amino acid protein sequence is set forth as follows: MAASNKEQSKLHIAMFPWFAYGHINPFIQLSNKLSSHGIQISFFSIPGNIDRIKSSLNLSPPN QLIPLTIPPTEGLSPNFDSSSEVTPQTAQLLTLALDQMQPQVKALFPHPQPQVILFDFAYHW LPSVASELGIKAVHFNTFPAVINSYLTVPSRMTDPNKPPTFEDLKNPPQGYPKTSTASVKTF EAQDYLFLFKSFDGGPCHFEKILAFTNSCDAILYRTCNEIEGPFIDYFKTQINKPLLLAGPSV PLPPSGELDEKWEMWLGKFPEKSVIYCSFGSETYLNDAQIQELTLGLELTGLPFILVLNFGT SNSTDAHNKLEASLPEGFRERIKDRGVLHTGWVQQQNILAHRSIGCFLTHAGFSSVIEGIV NDCQLAFLPLKADQFMIAKLFSGDLKAGVEVNRRDEDGSFAKEDIFEAIKTIMVDTDKEP SRSIRENHSNWRKFLMNKEIEASYIANLAHELKALVQKA (SEQ ID NO: 118).

[0056] Spinach GT2 nucleic acid gene sequence is set forth as follows: atgtgtgacgacaaaaaatcatctgttttgagcatagcattttatccgtggtttgctcttggtcaccttacttcatttctccgattagccaacaaacttgc acaaaatggtcacaatgtgtcctatttttatcccaactaatacattacctagattacttcctcacaaccattaccctggccaccttactttcatccccgtc accgtcccacccgttgacggcctccctctcagagccgagaccaccaacgatgtcccctcctcggctatacaccttcttatgactgccatggatttg acccgtgacactatcgaggcccatttggttagtatcaaacccgatgttgttttctacgactttgcttattggattcccgatctagcccgaaaacacgg

gttcaagtcagtactctacattacatcctatatagcaagatgtgcttattttgcccccgatttgaagtcgggtcatcagtccactggggccgaaattat tgcgccaccaccgggttttccgtctcagcatttccggatgcaagcacacgaggctgagactgtggcagacgtaggtaaagagcaagatggatt acaaggtataactatttctgaaaggatgcgcattgcttttggaaaatgcgacgcaattggagtaaagagttgtaaggagatggaaaaggtgtatat tgactactgtgagaagatatttggtaagtctgtactactagcaggtcctatggtccctaaaaccccatcttccaaacttgatgaatattttgatggttg gcttacgggttttggtgctgctactgtgatttattgtgcatttgggagtgaatgtgttctcgaaattaaccaatttcaacaacttcttcttggactagagc tcacaggaaggccatttttggtggccatgaagccgcctaagaagtatgaaacaatagagtcggccttaccagaagggtttgagaagagaacaa aaggaaggggaatcgtacatgagggttgggtgcagcaacaactgatattgcaacatccatcagtaggatgtttcataactcattgtggagttggg tctctttcggaagctatggtcagcaaatgtcaagtagtgttgatgcctcaagctgtagaccaattcatcaatgcgaggatgatgagtttagagttga agattggggttgaggttgagaagagagaagatgatggtttgttcacaaaggaggctgtgcataaggcggtctctttggtgatggaggaagaaa gtgaagtcgcaaaagagatgagggtaagtcatgataaatggagagaattcttattacaggaaggtcttgaggattcttatatcagtagcttcattca gagtctacgacagttaacgattggatga (SEQ ID NO: 119).

[0057] Spinach GT2 amino acid gene polypeptide is set forth as follows: MCDDKKSSVLSIAFYPWFALGHLTSFLRLANKLAQNGHNVSYFIPTNTLPRLLPHNHYPG HLTFIPVTVPPVDGLPLRAETTNDVPSSAIHLLMTAMDLTRDTIEAHLVSIKPDVVFYDFA YWIPDLARKHGFKSVLYITSYIARCAYFAPDLKSGHQSTGAEIIAPPPGFPSQHFRMQAHE AETVADVGKEQDGLQGITISERMRIAFGKCDAIGVKSCKEMEKVYIDYCEKIFGKSVLLA GPMVPKTPSSKLDEYFDGWLTGFGAATVIYCAFGSECVLEINQFQQLLLGLELTGRPFLV AMKPPKKYETIESALPEGFEKRTKGRGIVHEGWVQQQLILQHPSVGCFITHCGVGSLSEA MVSKCQVVLMPQAVDQFINARMMSLELKIGVEVEKREDDGLFTKEAVHKAVSLVMEEE SEVAKEMRVSHDKWREFLLQEGLEDSYISSFIQSLRQLTIG (SEQ ID NO: 120).

[0058] Values (±SD) represent mean of three independent biological experiments. Statistically significant differences compared with control plants (CHLH silenced) are indicated; *P < 0.05, **P < 0.01, ***P < 0.001.

[0059] **FIGURES 30A-30C** present data showing putative characterization of saponins accumulated in spinach with silenced (VIGS) glycosyltransferases. (**FIGURE 30A**) Relative content of selected compounds in plants with silenced SOAP6-SOAP9 (SEQ ID NO: 55, 57, 59, and 61, respectfully) and other two GTs. Values (±SD) represent mean of three independent biological experiments. Statistically significant differences compared with control plants (CHLH silenced) are indicated; **P < 0.01, ***P < 0.001. (**FIGURE 30B**) MS/MS spectra with the main fragments marked. (**FIGURE 30C**) Putative structures of accumulated saponins are based on accurate mass measurement and fragmentation patterns.

[0060] **FIGURES 31A-31F** present characterization of SOAP10 acetyltransferase in *planta* (spinach) and *in vitro*. (**FIGURE 31A**) LC-MS analysis of samples from spinach plants with VIGS silenced SOAP10 (red trace) or CHLH (black trace) as control. Blue and yellow regions of the

chromatogram contain desacetyl and acetylated saponins, respectively. Silencing of SOAP10 in spinach leaves resulted in decreased production of acetylated saponins and increased accumulation of desacetyl counterparts. (**FIGURE 31B**) Relative content of selected desacetyl saponins (Yossoside IV and XII) and their acetylated counterparts (Yossoside V and VII) in spinach plants with VIGS silenced acyltransferases and CHLH as control. The ordinate axis is in log10 scale. (**FIGURE 31C**) Ratio of desacetyl to acetylated saponins in VIGS silenced spinach plants. Silencing of SOAP10 resulted in increased accumulation of desacetyl saponins. (**FIGURES 31D, 31E, and 31F**) SOAP10 recombinant protein activity in vitro. EICs of desacetyl saponins (**FIGURE 31D** – m/z 1101.51; **FIGURE 31E** – m/z 1263.57; **FIGURE 31F** – m/z 1395.61) and acetylated counterparts (**FIGURE 31D** – m/z 1313.52; **FIGURE 31E** – m/z 1305.58; **FIGURE 31F** – m/z 3137.62) in: substrate (black trace) – purified fraction of desacetyl saponins from spinach; *in vitro* enzymatic assay with recombinant SOAP10 incubated with substrate in presence of acetyl-CoA (orange trace); and in extract of spinach leaves (green trace) for comparison. Dotted line separates regions on chromatogram with desacetyl and acetylated saponins on the left- and right-hand side, respectively. "$\Delta m/z = 42.01$" depicts difference in molecular weight of the compounds resulting from acetylation. (*SoAT1* (SEQ ID NO: 109 [gene] and SEQ ID NO: 113 [polypeptide]), *SoAT2* (SEQ ID NO: 110 [gene] and SEQ ID NO: 131 [polypeptide]), *SoAT3* (SEQ ID NO: 111 [gene] and SEQ ID NO: 115 [polypeptide]), and *SoAT4* (SEQ ID NO: 112 [gene] and SEQ ID NO: 116 [polypeptide]), for functional activities of *SoAT1-SoAT4* see Table 15.)

[0061] **FIGURES 32A-32D** present data from the expression of SOAPs 1-4 and 6-10 in *N. benthamiana*. (**FIGURE 32A**) EIC of medicagenic acid and its (poly)glycosylated derivatives [m/z = 501.32 (MA); m/z = 663.38 (MA+hex); m/z = 825.43 (MA+2hex); m/z = 987.48 (MA+3hex); m/z=1149.53 (MA+4hex) in negative ion mode] from plants transiently expressing either SOAP1-4 alone or all nine SOAPs without SOAP5 (SEQ ID NOs: 65 or 93 [gene], SEQ ID NO: 66 [polypeptide]), compared to control. Inset shows region of the chromatogram magnified 20 times. (**FIGURES 32B**, **32C**, and **32D**) show MSE spectra of mono-, bis- and trishexosylmedicagenic acid, respectively. Arrows represent losses of hexosyl moieties.

[0062] **FIGURES 33A-33F** present technical analysis of Cellulose Synthase Like G gene (SOAP5; SEQ ID NOs: 65 or 93 [gene], SEQ ID NO: 66 [polypeptide]). The SOAP5 encoded enzyme attaches glucuronic acid to the triterpenoid aglycone and plays a key role in spinach saponin biosynthetic pathway. (**FIGURE 33A**) Analysis of SOAP5 relative expression in spinach leaves after Virus Induced Gene Silencing (VIGs). Presented values were obtained from 4 independent biological replicates. Expression of SOAP5 was compared to the control (plants with silenced magnesium chelatase subunit H alone – CHLH). (**FIGURE 33B**) Medicagenic acid accumulation in spinach

leaves with silenced SOAP5 and other glycosyltransferases compared to control (CHLH). Significant differences are indicated, *** P<0.001. (**FIGURE 33C**) EIC of medicagenic acid 3-O-glucuronide [m/z = 677.35 (MA-3-GlcA), in negative ion mode] from plants transiently expressing SOAP1-5, compared to plants expressing SOAP1-4 alone and to control (empty vector - EV). (**FIGURE 33D**) EIC of MA-3-GlcA [m/z = 677.35 in negative ion mode] from yeast expressing SOAP1-5 + SoUGD1 compared to yeast expressing SOAP1-5, yeast controls (SOAP1-4+SoUGD1; SOAP5 alone and yeast cells transformed with EV). (**FIGURE 33E**) Schematic representation of SOAP5 activity in spinach. SOAP5 attaches glucuronic acid (highlighted in red) to medicagenic acid at position C-3. (**FIGURE 33F**) Spinach saponin biosynthetic pathway reconstitution in *N. benthamiana*. EIC of Yossoside V [m/z = 1305.57 in negative ion mode] in samples from *N. benthamiana* leaves transiently expressing all the ten SOAP genes compared to control plants (EV) and to spinach leaf extract.

[0063] **FIGURES 34A-34B** present the data showing the results of silencing SOAP5 in spinach (VIGS). (**FIGURE 34A**) EIC of medicagenic acid [m/z = 501.3, in negative ion mode] from plants with silenced SOAP5 (SEQ ID NOs: 65 or 93), compared to control (CHLH silenced) and to wild type (WT). Only chromatograms from *in planta* experiments are to scale. (**FIGURE 34B**) Tandem mass spectroscopy (MS/MS) of medicagenic acid [25V, positive ion mode] accumulated in the plant compared to the authentic standard. MA- medicagenic acid.

[0064] **FIGURES 35A-35C** present data showing transient expression of SOAP5 (SEQ ID NOS: 65 OR 93) in *N. benthamiana*. (**FIGURE 35A**) EIC of medicagenic acid and medicagenic acid 3-glucuronide [m/z = 501.32 (MA); m/z = 677.35 (MA-3-GlcA), in negative ion mode] from plants transiently expressing SOAPs1-5 (SEQ ID NOs: 45, 46, 51, 53, and 65, respectively), compared to plants expressing SOAPs1-4 alone (SEQ ID NOs: 45, 46, 51, and 53, respectively), SOAP5 alone (SEQ ID NOS: 65 OR 93) and to control (empty vectror-EV). Chromatograms are to scale. (**FIGURE 35B**) Structure of MA-3-GlcA with marked fragmentation routes (**FIGURE 35C**) MS/MS of MA-3-GlcA [45V, negative ion mode], arrow indicates loss of glucuronic acid moiety. (**FIGURE 35D**) shows important HMBC correlations of the hydrogens of medicagenic acid 3-*O*-glucuronide (MA-3-GlcA). Black arrows represent HMBC correlations of the six methyl groups of MA-3-GlcA. Red arrows represent HMBC correlations of other (selected) hydrogens of MA-3-GlcA. Numbers correspond to carbon atoms in medicagenic acid and glucuronic acid moiety. For more details see **Table 1**.

[0065] **FIGURES 36A-36F** present the substrate specificity of SOAP5 (SEQ ID NO: 66) (See **Table 16**). EIC of oleanolic acid 3-GlcA (**FIGURE 36A**), augustic acid 3-GlcA (**FIGURE 36B**), and gypsogenic acid 3-GlcA (**FIGURE 36C**) [m/z = 631.38 (OA-3-GlcA); m/z = 647.38 (AA-3-

GlcA); m/z = 661.35 (GA-3-GlcA), in negative ion mode] from plants transiently expressing SOAP1/2/5 (**FIGURE 36A**), SOAP1/2/3/5 (**FIGURE 36B**), SOAP1/2/4/5 (**FIGURE 36C**) compared to plants expressing same set of SOAPs without SOAP5 and to control (EV). Chromatograms are to scale. (**FIGURES 36D, 36E, and 36F** show tandem mass spectroscopy (MS/MS) of glucuronate derivatives from **FIGURES 36A, 36B, and 36C**, respectively, [40V, negative ion mode], arrows indicate loss of glucuronic acid (Am/z = 176.03; GlcA- H2O).

[0066] **FIGURES 37A-37F** show the products of SOAPs activity in *N. benthamiana*. EICs of Yossoside I (m/z = 823.42) (**FIGURE 37A**), Yossoside II (m/z = 969.27) (**FIGURE 37B**), Yossoside III (m/z = 1131.52) (**FIGURE 37C**), Yossoside IV (m/z = 1263.57) (**FIGURE 37D**) and Yossoside V (m/z = 1305.58) (**FIGURE 37E**) from plants expressing transiently combinations of SOAP genes indicated on each chromatogram. For MS/MS spectra see **FIGURE 40B**.

[0067] **FIGURES 38A-38E** shows mass spectroscopy (MS) based identification of the main products of SOAPs activity in *N. benthamiana*. (**FIGURE 38A**) MS/MS of Yossoside I (**Compound 7**)- product of SOAP6 activity on MA-3-GlcA. (**FIGURE 38B**) MS/MS of Yossoside II (**Compound 8**)- product of SOAP7 activity on Yossoside I. (**FIGURE 38C**) MS/MS of Yossoside III (**Compound 9**)- product of SOAP8 activity on Yossoside II. (**FIGURE 38D**) MS/MS of Yossoside IV (**Compound 10**)- product of SOAP9 activity on Yossoside III. (**FIGURE 38E**) MS/MS of Yossoside V (**Compound 11**)- product of SOAP10 activity on Yossoside IV.

[0068] **FIGURES 39A-39C** present the results of SOAP9 activity in *N. benthamiana*. (**FIGURE 39A**) EIC of medicagenic acid 3-GlcA-Xyl [m/z = 809.40 in negative ion mode] from plants transiently expressing SOAP1-5+SOAP9 compared to plant expressing SOAP1-5 alone and to control (EV). (See **Table 16**) (**FIGURE 39B**) MS/MS of 809.4 (45V), arrow indicates loss of glcA-xyl moiety (Am/z = 308.07). (**FIGURE 39C**) Structure of the SOAP9 product (**Compound 10**) with fragmentation patterns indicated by arrows.

[0069] **FIGURES 40A-40B** show the results of expression of SOAP5 in yeast. (**FIGURE 40A**) EIC of medicagenic acid and medicagenic acid 3-GlcA [m/z = 501.32 (MA); m/z = 677.35 (MA- 3-GlcA), in negative ion mode] from yeast expressing SOAP1-5 + UGD1 (SEQ ID NO: 74 [gene] and SEQ ID NO: 75 [protein]) compared to plants expressing SOAP1-5, yeast controls and to medicagenic acid authentic standard. Chromatograms are to scale. (**FIGURE 40B**) MS/MS of medicagenic acid 3-GlcA from yeast cells [45 V, negative ion mode], arrows indicate loss of glucuronic acid.

[0070] **FIGURES 41A-41F** show SOAP5 substrate specificity in yeast cells. EIC of oleanolic acid 3-GlcA (**FIGURE 41A**), gypsogenic acid 3-GlcA (**FIGURE 41B**) and Bayogenin 3-GlcA (**FIGURE 41C**) [m/z = 631.38 (OA-3-GlcA); m/z = 661.35 (GA-3-GlcA); m/z = 663.37 (Bayogenin-3-GlcA), in negative ion mode] from yeast expressing SOAP1-5 + UGD1 compared to

cells expressing SOAP1-4 + UGD alone, SOAP1-5 alone, SOAP1-4, SOAP1-2, SOAP5 alone, and to control (EV). Chromatograms are to scale. (**FIGURES 41D, 41E, and 41F**) MS/MS of glucuronate derivarives from A, B and C [45V, negative ion mode], arrows indicate loss of glucuronic acid (Am/z = 176.03; GlcA- H2O).

[0071] **FIGURES 42A- 42F** show Glucuronosyltransferase activity of Cellulose Synthase Like G (CSLG) enzymes is conserved among species belonging to phylogenetically distant orders. (**FIGURE 42A**) The phylogenetic tree depicts close evolutionary relationship of SOAP5 from spinach with some CSLG proteins from *Caryophyllales*, *Malvales*, *Apiales, Fabales, and Solanales*. The tree consists of proteins belonging to Cellulose Synthase (CESA) and several families of Cellulose Synthase Like enzymes (CSLA, CSLB, CSLE and **CSLG**). Dotted line encircles clades with CSLG proteins, while blown up region corresponds to CSLG closely related to spinach SOAP5 including GAME15 and CSLG in the *Fabales* ordera. (**FIGURE 42B**) Medicagenic acid is glucuronidated by CSLG from many species. EIC of MA-3-GlcA [m/z = 677.35 in negative ion mode] of samples from *N. benthamiana* leaves transiently expressing SOAP1-4 combined with CSLG proteins from spinach (SOAP5; SEQ ID NO: 66), *Chenopodium quinoa* (CqCSL; SEQ ID NO: 96), *Beta vulgaris* (BvCSL; SEQ ID NO: 94), *Medicago sativa* (MsCSL; SEQ ID NO: 98), *Glycine max* (GmCSL; SEQ ID NO: 100), *Glycyrrhiza uralensis* (GuCSL; SEQ ID NO: 81 or SEQ ID NO: 102) and *Lotus japonicus* (LjCSL; SEQ ID NO: 104), compared to plants expressing SOAP1-4 alone and to control (EV). (**FIGURE 42C**) Cellulose synthase like G proteins are key enzymes in the biosynthesis of triterpenoid saponins in multiple plant species. β-Amyrin (in the center) is oxidized by cytochromes P450 (blue circle) giving rise to multiple aglycons (1 – bayogenin; 2 – serjanic acid; 3 – oleanolic acid; 4 – medicagenic acid; 5 – glycyrrhetinic acid; 6 – soyasapogenol A; 7 – soyasapogenol B) that are decorated by CSLGs (pink circle) and other glycosyltransferases giving rise to triterpenoid saponins (8 – bayogenin -hexA-hex-hex (*M. sativa*); 9 – serjanic acid -hexA-hex (*Chenopodium quinoa (C. quinoa)*); 10 – betavulgaroside IV (*B. vulgaris*); 11 – yossoside V (*S. oleracea*); 12 – glycyrrhizin (*G. uralensis*); 13 – soyasapogenol A – hexA-hex-pent (*L. japonicus*); 14 – soyasaponin VI (*G. max*)). (**FIGURE 42D**) Glycyrrhizin biosynthetic pathway in *G. uralensis*. Highlighted functional groups are the result of the activity of specified enzymes. (**FIGURE 42E**) Extracted ion chromatogram (EIC) measurements of glycyrrhetinic acid 3-O-glucuronide and glycyrrhizin [m/z 645.36 and m/z 821.40 in negative ion mode] from *N. benthamiana* leaf samples transiently expressing five genes from glycyrrhizin biosynthetic pathway (bAS (SEQ ID NO: 45) + CYP88D6 (SEQ ID NO: 76) + CYP72A154 (SEQ ID NO: 78) + GuCSL (SEQ ID NO: 80 or SEQ ID NO: 103), and UGT73P12 (SEQ ID NO: 84)) compared to samples expressing four genes (-UGT73P12; (SEQ ID NO: 84)) and to control (bAS (SEQ ID NO: 45) + CYP88D6 (SEQ ID NO:

76) + CYP72A154 (SEQ ID NO: 78)). **(FIGURE 42F)** Presents a phylogenetic tree of Cellulose Synthase Like (CSL) polypeptides with special emphasis on Cellulose Synthaes Like G (CSLG). Blue background region consists of CSLG enzymes from many plant orders (i.e., *Carylophyllales*, *Fabales*, *Apiales*, *Malvales*, and *Solanales*). Within the CSLG enzymes (blue background), related subclades include the subclade drawn in red containing CSLG enzymes (in bold) with proven glucuronosyltransferase activity towards triterpenoid aglycones – CSLG from soybean, licorice, *Lotus japonicus*, alfalfa, red beet, quinoa, and spinach, and the subclade drawn in blue consisting of CSLGs from *Solanales* (*Solanum tuberosum*, potato; *Solanum lycopersicum*, tomato; *Solanum dulcamara*, bittersweet) and other orders; enzymes in bold are involved in steroidal alkaloid and steroidal saponin biosynthesis.

[0072] **FIGURES 43A-43C** shows LC-ESI-QTOF-MS analysis of saponins in quinoa, beetroot and alfalfa. Total Ion Chromatogram (TIC) of 80% methanolic extract of (**FIGURE 43A**) one-month old quinoa plants (*Chenopodium quinoa* var. Read Head), (**FIGURE 43B**) one-month old beetroot of (*Beta vulgaris* var. Bohan) and (**FIGURE 43C**) 2-week-old seedlings of alfalfa (*Medicago sativa*). Semitransparent rectangles mark chromatogram region with saponins. For putative characterization see **Table 12**.

[0073] **FIGURES 44A-44C** show virus induced gene silencing (VIGS) of CSLG in red beet (**FIGURE 44A**) Virus induced gene silencing in red beet; EV - empty vector, CHLH - magnesium chelatase subunit H, *BvCSL* - beetroot *cellulose synthase like G* (SEQ ID NO: 95 [gene]; VIGs silencing sequence SEQ ID NO: 107). (**FIGURE 44B**) Analysis of relative content (peak area) of triterpenoid saponins and oleanolic acid in red beet plants with silenced *BvCSL*. The ordinate axis is in $\log_{10}$ scale. Values represent mean of three independent biological experiments. Statistically significant differences compared with control plants (*CHLH* silenced alone) are indicated; *P < 0.05, ***P < 0.001. (**FIGURE 44C**) EIC of oleanolic acid [m/z = 455.35, negative ion mode] from plants with silenced *BvCSL* compared to control (*CHLH* and WT) and authentic standard. (For metabolite identification see **Table 12** *M. sativa*.)

[0074] **FIGURES 45A-45B** show silencing of CSLG in alfalfa hairyroots. (**FIGURE 45A**) EIC of representative saponins containing glucuronic acid attached to the aglycone from hairyroot cultures with silenced *Medicago sativa* Cellulose Synthase Like (MsCSL) compared to control (EV). (**FIGURE 45B**) Analysis of triterpenoid saponins in alfalfa hairyroots with silenced *MsCSL*. The ordinate axis is in $\log_{10}$ scale. Values represent mean of five independent biological experiments. Statistically significant differences compared with control (roots transformed with EV alone) are indicated; *P < 0.05, ***P < 0.01, ***P < 0.001. (asRNA – antisense RNA) The asRNA sequence used is set forth here:

CTACCCACTCTTCCGCTTCATAGATATAATCCCCTCCACAATAGGATGACTTAGAGCA
AGTATATAGATAATGAGGAAAAGTTGACCAAACATTTCATCAAAATCCCTCACATTG
AGTAGTCTCCATAAACCACCAAAGAAGCAAACAATATTTATTGTGAGTAACACAATC
ACTGGTGCCATGAGCAATGCTGCACCATCAAAATTAAACTTACCCTGCTCATATTTCT
TTAGCTTATCTTTGTCAATTGCTTTGTCCGATAAAGTGAATTTTGCCTTGTTCAACCCA
AACTTTTTTTTCGCTGTTTCTATAATTGCGAAAACGCACCCAATTGATTTTACAATCCC
ACTTCTTTGTTCATTCCACCAAATCATCGAGGAGCCACCAGTTGAAATTACCTCAATA
TAATGATGAATTTGACTGGATAGATACAATACTGTAAACACTATAAACCAAGGATCT
GTAACCTTTGGATATACAGGTATTCCCTTCAAGAAGCAAATCTGAGGAACGAATCCA
TATAAGATGTAGGCTGTTGCAAATTGGGTTGTGCTTACCAAGTAACAAAAAGTTAAA
CAATGAATAGTGGGCAATCTTGAAAGGCCATAAGTGAATGGGCTATATTTAGAGATT
GCAAGCAACAAAAGTTCAGATGACCACTTTATCGGTTGAATCAATCCCTCTTTCATGT
CAGTTGGAGCACATCCTAAGAAACATGGTCTTTTAGGGTAAAGATAAGTTGATCTCC
ATCCTCTACAATGGAGGAGATAGCCAGTAACGGTACTCTCAAGTTTTATAGCATACG
AGAATCCCACCTCATTACCCCAATTTGTGTTTCTCTCATAGGAACAAGAAGCCACTTC
ACATGCTTCTTGTAAAATTACATCTCTTGAAATATTCTGCTTCTTAGTTTGTTGACCAC
GTAACGCTTTTAGTGATTCTACATACATGTTAGACTTGCCAAAGTTGTATAGAGCATC
AAGTAGATAGTCCCCTTTTTGGTTTGGACTTCCAAAGAGTAATGCACTTCTACTTATA
TAATTTCCACTGCCAGTTAGACCTGGACCTCTTAATCCATCCATTCCCTTCCACTTTGT
CGTAAAAGCAGTCCTAGTCTGACTATCATATATGTCTTTCTTGCTAAGGTTGTGAAAC
ATTTGAGGGAATTGAACAAAAGCAACATCTTTAGAGGTTTCAGGATCAAGAAAAAAG
CACATGGATTGTTTGGCTGATGATGCATCATTACAATTCATATCACAATCTACGACAA
GTACATAAGGTCCATTGCTGATTAGCCCTGACACCCTAAGCAATGTATTGAGAGCTCC
TCCTTTGAATCTGTGAGGAACAGATGGTCTTTTTTCACGAGATACATAAACAACTAGT
GGCATTTCTTTTTGGTCATTAATGATCTCGATCCGAGAAGGTCTATCGGTCACCATAC
AAAGATTCTTTAGGTTGCTTCCGAATTTCTCAATATTTTTCTGCATTTTCTCGTATTTG
GCTTTGGTGGGCATATGTTCCTTGGAGTTTTCAACTTGTGATGAGTTAGGATCCACTCT
ATTTAATCTCTTCTTAATCTGGTCCCTCACTTCTTCAAATTCACGAGGTCGATGAAGTC
GTTCATTCTCACCCAAAGCAGTGAAGAAAAACTTAGGACACCTTGACTTAACATCAT
ATTTTTTACAAAAAGGAACCCAAACTTTAGCAAATTCAAAAGCCTCTTTGATCCCAAA
AAGAGTAATAGGAGAACCTCCATCATCAGAAAGATAAATAGAAAGTTTATTAGAAGG
GTAATCCATTGCAATAGCAGAAATAACAGTGTTCATAACATCAACCGTTGGTTCTTTT
TCAGGATCAATGGTACACACAAATATGTCGAGTCCCGGCAACTTCTCCTCCGGCGGTA
ATTTCTCGGTCATAACTGAACGATTCACCAGCCTCCAACGGAATGCTTGGTTGAAAAA

CCATAGAAATGATAGAATAATCTCAGCTATTGTCATTAGAAACCATGGATATGAAAT AAACAAATTGCTGATACGGTAGTAAAAGAGAAACAAGACACATGTGAAGTGGAAGA TTATGTAAGCTCTTCTTAGAGGTAACAATGGTTGAACTGTTTCTTTGTGAAATGTGAA GGTTGCCAT (SEQ ID NO: 108)

[0075] **FIGURES 46A-46E** show that the formation of glycyrrhetinic acid 3-O-monoglucuronide is catalyzed by GuCSLG. (**FIGURE 46A**) EIC of glycyrrhetinic acid 3-O-monoglucuronide [m/z = 645.36, in negative ion mode] of samples from *N. benthamiana* expressing proteins from *Glycyrrhiza uralensis*: bAS - β-amyrin synthase (SEQ ID NO: 48); CYP88D6 - β-amyrin 11-oxidase (SEQ ID NO: 77); CYP72A154 - 11-oxo-b-amyrin 30-oxidase (SEQ ID NO: 79); *GuUGAT* (SEQ ID NO: 83)- glycosyltransferase previously reported to perform glucuronation of glycyrrhetinic acid; UGT73P12 (SEQ ID NO: 85)- glycosyltransferase transferring glcA onto glycyrrhetinic acid 3-O-monoglucuronide; *GuCSL*- cellulose synthase like G. (**FIGURE 46B**) MS/MS of 654.36 (glycyrrhetinic acid 3-O-monoglucuronide isomer 1; 55 V). (**FIGURE 46C**) MS/MS of 654.36 (isomer 2; 55 V). (**FIGURE 46D**) Structure of the glycyrrhetinic acid 3-O- monoglucuronide with fragmentation patterns indicated by arrows. (**FIGURE 46E**) Nomenclature of the cross-ring fragmentation of glucuronic acid with detected ions.

[0076] **FIGURES 47A-47F** show formation of glycyrrhizin (glycyrrhetinic acid 3-O-diglucuronide) in *N. benthamiana* requires GuCSLG. (**FIGURE 47A**) EIC of glycyrrhizin [m/z = 821.40, in negative ion mode] of samples from *N. benthamiana* expressing proteins from *G. uralensis*. bAS - b-amyrin synthase [β-amyrin synthase] (SEQ ID NO: 48); CYP88D6 - b-amyrin 11-oxidase (SEQ ID NO: 76 [gene] and SEQ ID NO: 77 [polypeptide]); CYP72A154 - 11-oxo-b-amyrin 47-oxidase (SEQ ID NO: 78 [gene] and SEQ ID NO: 79 [polypeptide]); *GuUGAT* (SEQ ID NO: 82 [gene] and SEQ ID NO: 83 [polypeptide])- glycosyltransferase previously reported to perform glucuronation of glycyrrhetinic acid; UGT73P12 - glycosyltransferase transferring glcA onto glycyrrhetinic acid 3-O-monoglucuronide (SEQ ID NO: 84 [gene] and SEQ ID NO: 85 [polypeptide]); *GuCSL* - cellulose synthase like G (SEQ ID NO: 80 or SEQ ID NO: 103 [gene], and SEQ ID NO: 81 or SEQ ID NO: 102 [polypeptide]). (**FIGURE 47B**) MS/MS (65 V) of 821.40 (isomer 1) - product of bAS (SEQ ID NO: 45)+ CYP88D6 (SEQ ID NO: 77) + CYP72A154 (SEQ ID NO: 79)+ *GuCSL* (SEQ ID NO: 81 or SEQ ID NO: 102)+ UGT73P12 (SEQ ID NO: 85). (**FIGURE 47C**) MS/MS (65 V) of 821.40 (isomer 2) - product of bAS (SEQ ID NO: 45)+ CYP88D6 (SEQ ID NO: 77) + CYP72A154 (SEQ ID NO: 79)+ *GuCSL* (SEQ ID NO: 81 or SEQ ID NO: 102)+ UGT73P12 (SEQ ID NO: 85). (**FIGURE 47D**) and (**FIGURE 47E**) MS/MS of glycyrrhizin authentic standard, isomer 1 and 2 respectively. (**FIGURE 47F**) Structure of glycyrrhizin with fragmentation patterns indicated by arrows.

[0077] **FIGURES 48** presents the structure of SOAP5. (**FIGURE 48**) Alignment of protein sequences of Arabidopsis AtCESA1 (SEQ ID NO: 67), AtCESA3 (SEQ ID NO: 68), and their orthologs in spinach (*SoCESA1* (SEQ ID NO: 70) and *SoCESA3* (SEQ ID NO: 71)) together with SOAP5 (SEQ ID NO: 66) and its counterpart in Arabidopsis – *AtCSLGl* (SEQ ID NO: 69).

[0078] **FIGURES 49A-49C** present the predicted 3D structure of SOAP5. (**FIGURES 49A and 49B**) 3D model of SOAP5 with predicted tunnels, allowing medicagenic acid to reach the active site of the enzyme (blue - 1,2,3) and to leave (yellow - 4) the hydrophobic environment of the lipid bilayer. (**FIGURE 49C**) Profiles and properties of the tunnels shown in (**FIGURE 49A**).

[0079] **FIGURES 50A and 50B** present site directed mutagenesis data of SOAP5. (**FIGURE 50A**) EICs of medicagenic acid [m/z = 501.32] and medicagenic acid 3-GlcA [m/z = 677.35] from plants transiently expressing SOAP1-4 with mutated versions of SOAP5 compared to native SOAP5, SOAP1-4 alone and control (EV). (**FIGURE 50B**) Accumulation of MA and MA-3-GlcA in plants expressing mutated variants of SOAP5 represented as a ratio of MA-3-GlcA to free MA. The ordinate axis is in $\log_{10}$ scale. Values represent mean of three independent biological experiments. Statistically significant differences compared with plants expressing native SOAP5 are indicated; ***P < 0.001.

[0080] **FIGURES 51A – 51H** present micrographs showing that SOAP proteins including SOAP5 are co-localized in the endoplasmic reticulum (ER) membrane. (**FIGURES 51A-51F**) Confocal images of SOAP5:RFP transiently co-expressed in *N. benthamiana* leaves with (**FIGURE 51A**) ER:GFP marker; (**FIGURE 51B**) Golgi:GFP marker; (**FIGURE 51C**) SOAP1:GFP; (**FIGURE 51D**) SOAP2:GFP; (**FIGURE 51E**) SOAP3:GFP and SOAP4:GFP (**FIGURE 51F**). Each panel contains images of GFP, RFP and both channels merged. Scale bar- 50 μm. (**FIGURE 51G**) Topology model of SOAP5 with conserved motifs (red), substituted amino acids (blue, pointed with arrows) and six transmembrane domains indicated. (**FIGURE 51H**) FRET index of cells that expressed the indicated constructs. The values are means ± SE. Significant differences comparing to negative control expressing free GFP and RFP (NC) are indicated; * P<0.05; ** P<0.01.

[0081] **FIGURES 52A – 52F** present micrographs showing SOAPs involved in triterpenoid aglycon formation are colocalized and reside in endoplasmic reticulum (ER). (**FIGURES 52A-52F**) Confocal images of SOAPs in fusion with fluorescent proteins (GFP and RFP) transiently co-expressed in *N. benthamiana* leaves. (**FIGURE 52A**) SOAP1:GFP + SOAP2:RFP; (**FIGURE 52B**) SOAP1:GFP + SOAP3:RFP; (**FIGURE 52C**) SOAP1:GFP + SOAP4:RFP; (**FIGURE 52D**) SOAP2:GFP + SOAP3:RFP; (**FIGURE 52E**) SOAP2-GFP + SOAP4:RFP; (**FIGURE 35F**) SOAP3:GFP + SOAP4:RFP (F). Each panel contains images of GFP, RFP and both channels merged together. Scale bar - 50 picometers.

[0082] **FIGURES 53** present data showing that SOAPs in fusion with fluorescent proteins (FP) are

still active. EICs of medicagenic acid [m/z = 501.32] and medicagenic acid 3-GlcA [m/z = 677.35] from *N. benthamiana* plants transiently expressing SOAP1:eGFP, SOAP2:eGFP, SOAP3:mRFP, SOAP4:eGFP and SOAP5:mRFP compared to native SOAP5, SOAP1-4 alone and control (EV). *SOAP1:GFP, SOAP2:GFP, SOAP3:RFP, SOAP4:GFP* and *SOAP5:RFP* compared to native *SOAP1-5*, *SOAP1-4* alone and control (EV).

[0083] **FIGURES 54A-54B** present SOAPs in fusion with fluorescent proteins (FP) are still active. (**FIGURE 54A**) Fluorescence resonance energy transfer (FRET) index of cells that expressed the indicated constructs. The values are means ± SE. Significant differences comparing to negative control expressing free GFP and RFP (NC) are indicated with asterisks; * $P<0.05$; ** $P<0.01$. (**FIGURE 54B**) Schematic representation of possible protein localization within ER.

DETAILED DESCRIPTION

[0084] In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by those skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known methods, procedures, and components have not been described in detail so as not to obscure the present invention.

[0085] Specifically disclosed herein are genetically modified cells having increased expression of at least one heterologous gene compared to a corresponding unmodified cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said genetically modified cell comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified cell. In some embodiments, the genetically modified cell expressing a heterologous CSLG further expresses at least one additional heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, and an UDP-glucose 6-dehydrogenase 1, or any combination thereof.

[0086] The term "CSLG" encompasses Cellulose Synthase Like enzymes from the G family. There are various cellulose synthase Like (CSL) families, each with a different suffix, for example, A, B, E, and G. The suffix for each family is assigned according to enzyme's similarity to previously characterized proteins from *Arabidopsis thaliana.* CSL enzymes from different families have different technical characteristics. As used throughout, the terms "CSL" without a suffix of A, B, or

E, in certain embodiments, may be used interchangeably with the term "CSLG", having all the same meaning and qualities.

[0087] "CSLG" proteins, in certain embodiments, may encompass "GAME15" proteins and "SOAP5" proteins.

[0088] Described herein are six tested CSLG genes, which in certain embodiments include the GAME15 genes and their encoded enzymes (6 enzymes from 6 species – tomato, wild tomato, potato, wild potato, eggplant, and pepper), which transfer Glucoronic Acid or other sugar moieties to generate steroidal alkaloids and which transfer Glucoronic Acid or other sugar moieties on furostanol-type saponin (aglycone) to generate steroidal saponins or precursors thereof or to generate steroidal alkaloids or precursors thereof (see, e.g., **FIGURE 1**).

[0089] Described herein are seven tested CSLG genes (also known as SOAP5 genes) and their encoded enzymes (7 enzymes from 7 species - spinach, Chinese licorice, red beet, quinoa, alfalfa, soybean, and *Lotus japonicum*), which transfer glucuronic acid on triterpenoid aglycone. The term "triterpenoid aglycone" may in some embodiments, encompass the group of compounds derived from beta-amyrin. Thus, CSLG enzymes disclosed herein provide a unique activity.

[0090] A skilled artisan would appreciate that altering the content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a cell, for example a plant cell, a yeast cell, an algal cell, a bacterial cell, may comprise mutating an endogenous CSLG gene or regulating the CSLG gene expression, or may comprise adding a heterologous CSLG to a cell that does not express a CSLG enzyme or in order to supplement the activity of a CSLG enzyme expressed in the heterologous cell. In some embodiments, the plant cell may be comprised within a plant or plant part.

[0091] According to certain embodiments, expression of the at least one gene or any combination thereof is altered, the altering comprising mutagenizing the at least one gene, wherein the mutagenesis comprises introduction of one or more point mutations, or genome editing, or use or a bacterial CRISPR/CAS system, or a combination thereof. In certain embodiments, expression of the CSLG gene or polynucleotide is silenced, repressed, or reduced, e.g., by deletion, insertion or modification, or by introduction of at least one silencing molecule targeted to a CSLG gene.

[0092] It is to be understood that increasing the expression of the at least one gene or combination thereof may be achieved by various means, all of which are explicitly encompassed within the scope of present invention. According to certain embodiments, enhancing the expression of CSLG can be affected at the genomic and/or the transcript level using a variety of molecules that enhance transcription and/or translation including, but not limited to, transcription factors. Inserting a mutation to the at least one gene, including addition of promoters, enhancers, and the like can be also used, as long as the mutation results in up-regulation of the gene expression. According to other embodiments,

expression is increased at the protein level using agonists and the like.

[0093] In some embodiments, these genetically modified cells produce steroidal alkaloids. In some embodiments, these genetically modified cells produce steroidal saponins. In some embodiments, these genetically modified cells produce triterpenoid saponins. In some embodiments, the genetically modified cells comprise an increased content of at least a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell. In some embodiments, the genetically modified cells comprise an increased content of at least a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell. In some embodiments, the genetically modified cells comprise an increased content of at least a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell.

[0094] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[0095] In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[0096] In some embodiemtns, a CSLG enzyme homolog has at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 98%, or 99%identity of the aminio acid seuqence. In some embodiemtns, a gene encoding a CSLG enzyme homolog has at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 98%, or 99% identity of the nucleic acid seuqence.

[0097] In some embodiemtns, a CSLG enzyme homolog has at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 98%, 99%, or 100% coverage of the aminio acid seuqence. In some embodiemtns, a gene encoding a CSLG enzyme homolog has at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% 98%, 99%, or 100% coverage of the nucleic acid seuqence.

[0098] In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the

nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[0099] In some embodiments, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104. In some embodiments, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

[00100] In some embodiments, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

[00101] In some embodiments, the nucleic acid sequence encoding said at least one additional heterologous gene encodes: (a) a β-amyrin synthase, said nucleic acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or (b) a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a (c) glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or (d) an acetyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or (e) a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or (f) any combination thereof of (a), (b), (c), (d), and (e).

[00102] In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes: (a) a β-amyrin synthase, said amino acid sequence set forth in SEQ ID NO: 48; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 48; or (b) a cytochrome P450, said amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or

54; or a (c) glycosyl transferase, said amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or (d) an acetyltransferase, said amino acid sequence set forth in SEQ ID NO: 64; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 64; or (e) a UDP-glucose 6-dehydrogenase 1, said amino acid sequence set forth in SEQ ID NO: 75; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 75; or (f) any combination thereof of (a), (b), (c), (d), and (e).

[00103] Heterologous expression encompasses the expression of a gene or part of a gene in a host cell or host organism, for example but not limited to a plant cell, a yeast cell, an alga, or a plant, which does not naturally have this gene or gene fragment. Insertion of the gene in the heterologous host may be performed by any recombinant DNA technology known in the art, or for example in the case of yeast, may be performed by mating.

[00104] The triterpenoid saponin or said derivative, metabolite, or biosynthetic intermediate thereof, may have a broad range of commercial uses and may in some embodiments comprise a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof.

[00105] In some embodiments, (a) said steroidal glycoalkaloid comprises alpha-tomatine, tomatine, dehydrotomatine, hydroxytomatine, acetoxytomatine, dihydroxytomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, and their derivatives, related compounds or any combination thereof; (b) said steroidal saponin comprises _uttroside B, tomatosides and their derivatives, related compounds, or any combination thereof; (c) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), glycyrrhetinic acid 3-O-monoglucuronide (Compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; (d) the biosynthetic intermediate of said triterpenoid saponin comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or (e) any combination of these.

[00106] According to certain exemplary embodiments, the downstream steroidal glycoalkaloid is selected from the group consisting of esculeosides or dehydroesculeosides.

[00107] In some embodiments, the genetically modified cell further comprises an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00108] In some embodiments, an "altered content" encompasses an increased content of a phytosterol/a phytocholesterol/phytocholestenol, derivative thereof, metabolike thereof, or biosynthetic intermediate thereof. In some embodiments, an "altered content" encompasses an increased content of a phytosterol, derivative thereof, metabolike thereof, or biosynthetic intermediate thereof.

[00109] In some embodiments, the genetically modified cell further comprises a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00110] In some embodiments, the genetically modified cells are plant cells, yeast cells, or alga cells. In some embodiments, the genetically modified plant cells are comprised within a plant or a plant part. In some embodiments, the genetically modified cells are yeast cells. In some embodiments, the genetically modified cells are alga cells.

[00111] In some embodiments, said plant cell comprises a cell from a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

[00112] In some embodiments, (a) when said plant cell comprises a cell from a plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus; (b) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus; (c) when said plant cell comprises a cell from a plant in the

Solanales order, said plant is selected from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus; (d) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus; (e) when said plant cell comprises a cell from a plant in the *Malvales* order, said plant is selected from the *Theobroma* genus; (f) when said plant cell comprises a cell from a plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium* genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus; or (g) when said plant cell comprises a cell from a plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

[00113] In some embodiments, (a) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group consisting of spinach, beetroot, and quinoa; (b) when said plant cell comprises a cell from a plant in the *Solanales* order, said plant is selected from the group consisting of tomato, wild tomato, potato, wild potato, eggplant, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, ground cherry, tobacco, and bittersweet; or (c) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

[00114] In some embodiments, the plant cell is from a tomato plant having: (i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00115] In some embodiments, the plant cell is from a tomato plant having: (i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic

intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00116] In some embodiments, the plant cell is from a potato plant having: (i) an increased content of alpha-chaconine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00117] In some embodiments, the plant cell is from a potato plant having: (i) an increased content of alpha-chaconine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00118] In some embodiments, the plant cell is from an eggplant plant having: (i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative

thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00119] In some embodiments, the plant cell is from an eggplant plant having: (i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00120] In some embodiments, said plant cell comprises a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a callus cell, a bract cell, or a flower cell. In some embodiments, said plant cell is comprised in a plant or a portion thereof, said portion thereof comprising a plant leaf, a plant petiole, a plant stem or stalk, a plant root, a plant bud, a plant tuber, a plant bean, a plant grain or kernel, a plant fruit, a plant nut, a plant legume, a plant seed, a plant bract, or a plant flower.

[00121] In some embodiments, said yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus.

[00122] In some embodiments, said alga is selected from a microalga, a multi-cellular alga, a cyanobacterium, a diatom, chlorophytes (green algae), rhodophytes (red algae), or phaeo-phytes (brown algae), a *Dunaliella*, a *Chlamydamonas*, or a *Hematococus*.

[00123] Specifically disclosed herein are genetically modified plants comprising at least one cell having altered expression of at least a cellulose synthase like G (CSLG) gene compared to the expression of CSLG in a corresponding unmodified plant, and wherein the genetically modified plant has an altered content of at least one steroidal alkaloid, a derivative thereof a metabolite thereof; of at least one steroidal saponin , a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[00124] According to certain embodiments, expression of the at least one gene or any combination thereof is altered, the altering comprising mutagenizing the at least one gene, wherein the mutagenesis comprises introduction of one or more point mutations, or genome editing, or use of a bacterial CRISPR/CAS system, or a combination thereof.

[00125] According to certain embodiments, expression of the gene encoding at least one CSLG is elevated compared to its expression in the corresponding unmodified cell or plant, respectively. According to certain embodiments, the genetically modified cell or genetically modified plant comprises a polynucleotide encoding a CSLG, wherein expression of the polynucleotide is selectively increased. According to certain embodiments, the genetically modified plant comprises at least one cell comprising at least one transcribable polynucleotide encoding at least one protein selected from the group consisting of at least one CSLG.

[00126] According to certain embodiments, the genetically modified plant is a transgenic plant comprising at least one cell comprising at least one silencing molecule targeted to a polynucleotide encoding at least one CSLG. According to certain embodiments, the transgenic plant comprises a polynucleotide encoding a CSLG, wherein expression of the polynucleotide is selectively silenced, repressed, or reduced. According to certain embodiments, the transgenic plant comprises a polynucleotide encoding a CSLG, wherein the polynucleotide has been selectively edited by deletion, insertion, or modification to silence, repress, or reduce expression thereof, or wherein the genetically modified plant is a progeny of the gene edited plant. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to a CSLG gene. According to certain embodiments, the silencing molecule is selected from the group consisting RNA interference molecule and an antisense molecule, or wherein the silencing molecule is a component of a viral induced gene silencing system. According to certain embodiments, the silencing molecule comprises a polynucleotide having a nucleic acid sequence substantially complementary to a region of a CSLG gene.

[00127] According to some embodiments, the transgenic plant comprises a plurality of cells comprising the silencing molecule targeted to the at least one CSLG gene. According to additional embodiments, the majority of the plant cells comprise the silencing molecule.

[00128] It is to be understood that inhibiting the expression of the at least one gene or combination thereof may be achieved by various means, all of which are explicitly encompassed within the scope of present invention. According to certain embodiments, inhibiting the expression of CSLG can be effected at the genomic and/or the transcript level using a variety of molecules that interfere with transcription and/or translation including, but not limited to, antisense, siRNA, Ribozyme, or DNAzyme molecules. Inserting a mutation to the at least one gene, including deletions, insertions, site specific mutations, zinc-finger nucleases and the like can be also used, as long as the mutation results in down-regulation of the gene expression. According to other embodiments, expression is inhibited at the protein level using antagonists, enzymes that cleave the polypeptide and the like.

[00129] In some embodiments, said CSLG gene comprises the endogenous CSLG gene.

[00130] In some embodiments, said endogenous CSLG gene comprises a mutation, said mutation comprising at least one or more point mutations, or an insertion, or a deletion, or any combination thereof, and: (a) wherein said expressed CSLG enzyme has increased stability or increased activity or both and the altered content comprises an increased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant; or (b) wherein said expressed CSLG enzyme has decreased stability or decreased activity or both and the altered content comprises a decreased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[00131] In some embodiments, (a) said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression and said altered content comprises a reduced amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant; or (b) said endogenous CSLG gene is overexpressed and said altered content comprises an increased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[00132] In some embodiments, when said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression, said cell further comprises at least one silencing molecule targeted to the polynucleotide encoding said CSLG gene, wherein the silencing molecule is selected from an RNA interference molecule or an antisense molecule, or wherein the silencing molecule is a component of a viral induced gene silencing system.

[00133] In some embodiments, said CSLG gene comprises a heterologous CSLG gene, and said altered expression comprises a de novo expression of said gene.

[00134] In some embodiments, said expression comprises increased expression compared to a

corresponding unmodified cell, and said altered content comprising increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00135] In some embodiments, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40 , 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40 , 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00136] In some embodiments, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40 , 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00137] In some embodiments, the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104. In some embodiments, the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

[00138] In some embodiments, the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

[00139] In some embodiments, the altered expression of the at least one cellulose synthase like G (CSLG) gene is altered by introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, any combination thereof, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence.

[00140] In some embodiments, the altered expression of the at least one CSLG gene is altered by introducing a silencing molecule targeted to the at least one CSLG gene and wherein: (a) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 30, the silencing molecule is set forth in SEQ ID NO: 42 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 42 or a complementary sequence thereof; (b) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 34, the silencing molecule is set forth in SEQ ID NO: 43 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 43 or a complementary sequence thereof; (c) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 38, the silencing molecule is set forth in SEQ ID NO: 44 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 44 or a complementary sequence thereof; (d) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 65 or SEQ ID NO: 93, the silencing molecule is set forth in SEQ ID NO: 106 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 65 or the nucleic acid sequence set forth in SEQ ID NO: 93; or (e) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 95, the silencing molecule is set forth in SEQ ID NO: 107 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 95.

[00141] The triterpenoid saponin or said derivative, metabolite, or biosynthetic intermediate thereof, may have a broad range of commercial uses and may in some embodiments comprise a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof.

[00142] In some embodiments, (a) said steroidal glycoalkaloid comprises alpha-tomatine, tomatine, dehydrotomatine, hydroxytomatine, acetoxytomatine, di-hydroxytomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or their derivatives and pathway intermediates or any combination thereof; (b) said steroidal saponin comprises _uttroside B, tomatosides, and all related compounds, or their derivatives and pathway intermediates , or any combination thereof; (c) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), glycyrrhetinic acid 3-O-monoglucuronide (Compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26),

serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; (d) the biosynthetic intermediate of said triterpenoid saponin comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or (e) any combination of these.

[00143] In some embodiments, the genetically modified plant has an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00144] In some embodiments, the genetically modified plant has an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, wherein: (a) when the altered content of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof comprises an increased amount compared to the corresponding unmodified plant, the altered content of the phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or the phytocholestenol, or comprises a reduced amount compared to the corresponding unmodified plant a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (b) when the altered content of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin , a derivative

thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof comprises a reduced amount compared to the corresponding unmodified plant, the altered content of the phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; the cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or the phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof comprises an increased amount compared to the corresponding unmodified plant.

[00145] In certain embodiments, an increase in each or sterodial saponins, steroidal alakoids, and or triterpenoid saponins may cause a change in the content level of the other(s). In some emobdiments, the change comprises an increased content. In some embodiments, the change comprises a decrease content. In some embodiments, the change comprises an increased content of one and a decrease content of another.

[00146] In some embodiments, said plant comprises a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, or the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

[00147] In some embodiments, (a) when said plant comprises a plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus; (b) when said plant comprises a plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus; (c) when said plant comprises a plant in the *Solanales* order, said plant is selected from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus; (d) when said plant comprises a plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus; (e) when said plant comprises a plant in the *Malvales* order, said plant is selected from the *Theobroma* genus; (f) when said plant comprises a plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium* genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus; or (g) when said plant comprises a plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

[00148] In some embodiments, (a) when said plant comprises a plant in the *Caryophyllales* order, said plant is selected from the group consisting of spinach, beetroot, and quinoa; (b) when said plant comprises a plant in the *Solanales* order, said plant is selected from the group consisting of tomato, wild tomato, potato, wild potato, eggplant, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, ground cherry, tobacco, and bittersweet; or (c) when said plant comprises a plant in the *Fabales* order, said plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

[00149] In some embodiments, the plant is a tomato plant having: (i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00150] In some embodiments, the plant is a tomato plant having: (i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00151] In some embodiments, the plant is a potato plant having: (i) an increased content of alpha-chaconine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic

intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00152] In some embodiments, the plant is a potato plant having: (i) an increased content of alpha-chaconine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00153] In some embodiments, the plant is an eggplant plant having: (i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00154] In some embodiments, the plant is an eggplant plant having: (i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or (ii) a reduced content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00155] In some embodiments, said plant cell comprises a leaf cell, a petiole cell, a plant stem or

stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a callus cells, a bract cell, a callus cell, and a flower cell.

[00156] Specifically disclosed herein are methods of producing steroidal alkaloids, steroidal saponins, or triterpenoid saponins in a genetically modified cell, the methods comprising: (a) introducing an at least one heterologous gene into said cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said heterologous gene is optionally comprised in a vector; and (b) expressing said at least one heterologous gene in said cell; wherein said cell comprises an increased content of at least one steroidal alkaloid, at least one steroidal saponin, or at least one triterpenoid saponin compared to a corresponding unmodified cell.

[00157] In some embodiments, said introducing further comprises introducing an at least one additional heterologous gene into said cell, said heterologous gene selected from the group consisting of the group encoding a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, and a UDP-glucose 6-dehydrogenase I, or any combination thereof, wherein said at least one additional heterologous gene is optionally comprised in a vector; and further comprising expressing said at least one additional heterologous gene in said cell.

[00158] In some embodiments, (a) said at least one heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or (b) said at least one additional heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or (c)a combination thereof of (a) and (b).

[00159] In some embodiments, said introducing comprises transforming said at least one cell with (a) said at least one heterologous gene or a polynucleotide sequence encoding said at least one heterologous gene, or the vector comprising said at least one heterologous gene; or (b) said at least one additional heterologous gene or a polynucleotide sequences encoding said at least one additional heterologous gene, or the vector comprising said at least one additional heterologous gene; or (c) a combination thereof of (a) and (b); wherein said expressing comprises transient expression or constitutive expression.

[00160] Specifically disclosed herein are methods of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or a plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising: (a) transforming said at least one plant cell with at least one silencing molecule targeted to a nucleic acid gene sequence encoding a Cellulose Synthase Like G (CSLG) enzyme; or (b) mutagenizing at least one nucleic acid

sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; wherein expression of the gene encoding the CSLG enzyme is reduced in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell, wherein the plant comprising said genetically modified cell comprises reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00161] In some embodiments: (a) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 30, the silencing molecule is set forth in SEQ ID NO: 42 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 42 or a complementary sequence thereof; (b) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 34, the silencing molecule is set forth in SEQ ID NO: 43 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 43 or a complementary sequence thereof; (c)when the nucleic acid of the CSLG is set forth in SEQ ID NO: 38, the silencing molecule is set forth in SEQ ID NO: 44 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 44 or a complementary sequence thereof; (d) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 65 or SEQ ID NO: 93, the silencing molecule is set forth in SEQ ID NO: 106 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 65 or the nucleic acid sequence set forth in SEQ ID NO: 93; or (e) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 95, the silencing molecule is set forth in SEQ ID NO: 107 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 95.

[00162] In some embodiments: (a) said steroidal alkaloid, said steroidal saponin, or said triterpenoid saponin comprises a toxin or a bitter tasting compound or having hormone mimicking properties, or a combination thereof; or (b) said steroidal glycoalkaloid comprises alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or (c)said steroidal saponin comprises utteroside B and tonatosides and related compounds

(all steroidal saponins family), pathway intermediates and derivatives , or any combination thereof; or (d) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or (e)said intermediate comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or (f) a combination thereof.

[00163] In certain embodiments, reduction of a triterpenoid saponin is desired, for example, but not limited to triterpenoid saponins having a bitter taste, as in quinoa, or being toxic. Other examples of triterpenoid saponins wherein reduction may be desired include triterpenoid saponins that have toxic properties or mimic hormones. Thus, in certain embodiments, increased production of triterpenoid saponins is beneficial and in other embodiments, decreased production of triterpenoid saponins is beneficial.

[00164] In some embodiments, the method further comprises altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00165] In some embodiments, the method further comprises increasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00166] According to some embodiments, the method further comprises purifying the phytosterol, or derivative, metabolite, or biosynthetic intermediate thereof; the cholesterol, or derivative, metabolite, or biosynthetic intermediate thereof; the phytocholesterol, or derivative, metabolite, or

biosynthetic intermediate thereof; the cholestenol, or derivative, metabolite, or biosynthetic intermediate; or the phytocholestenol, or derivative, metabolite, or biosynthetic intermediate thereof, extracted from the transformed plant. According to certain embodiments, the phytosterol comprises phytocholesterol. According to some embodiments, the phytosterol, or derivative, metabolite, or biosynthetic intermediate thereof; the cholesterol, or derivative, metabolite, or biosynthetic intermediate thereof; the phytocholesterol, or derivative, metabolite, or biosynthetic intermediate thereof; the cholestenol, or derivative, metabolite, or biosynthetic intermediate; or the phytocholestenol, or derivative, metabolite, or biosynthetic intermediate thereof, extracted from the transformed plant, comprises a nutrient or food additive, a high-value steroidal compound (e.g., pro-vitamin D and/or diosgenin), an anti-cholesterol agent (e.g., a plant phytosterol competing with dietary cholesterol in a mammalian or avian gut), or a cosmetic agent.

[00167] Specifically disclosed herein are methods of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin , a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising: (a) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; and (b) expressing said nucleic acid encoding said CSLG; wherein the plant comprising said genetically modified cell comprises increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00168] In some embodiments: (a) said steroidal alkaloid, said steroidal saponin, or said triterpenoid saponin comprises a toxin or a bitter tasting compound or having hormone mimicking properties, or a combination thereof; or (b) said steroidal glycoalkaloid comprises alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or (c)said steroidal saponin comprises uttroside B, a tomatoside,, or any combination thereof; or (d) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin

(Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or (e)said intermediate comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or (f) a combination thereof.

[00169] In some embodiments: (a) expression of the gene encoding the CSLG enzyme is increased in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell; or (b) said encoded CSLG enzyme has increased activity in the genetically modified plant cell compared to its activity in a corresponding unmodified plant cell; or (c) said encoded CSLG enzyme has increased stability in the genetically modified plant cell compared to its stability in a corresponding unmodified plant cell; or (d) any combination thereof of (a), (b), and (c).

[00170] In some embodiments, the method further comprises altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00171] In some embodiments, the method further comprises reducing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00172] Each possibility described herein represents a separate embodiment of the invention.

Genetically Modified Cells & Genetically Modified Plants

[00173] Disclosed herein are genetically modified cells and genetically modified plants, wherein expression of key genes in the steroidal glycoalkaloid/steroidal saponin/triterpenoid saponin metabolic pathway (biosynthesis pathway of steroidal alkaloids, steroidal saponins, and triterpenoid saponins) has been altered. Altering the expression of these genes results in concomitant alteration in the steroidal alkaloid profile, in the steroidal saponin profile, or in the triterpenoid profile. In an non-

limiting example, changing the production level of steroidal alkaloid can result in improved plants comprising elevated content of steroidal alkaloids having increased resistance to pathogens, or plants having a reduced content of these secondary compounds in the plant edible parts and thus producing improved crops, wherein the improved crop has reduced or eliminated anti-nutritional content. Alternatively, or additionally, in another non-limiting example, controlling the expression of genes disclosed herein may be used for the production of desired steroidal alkaloids or plant-based cholesterol for further use, e.g., in the nutritional, cosmetic, or pharmaceutical industry. In another non-limiting example, disclosed herein are the means and methods for producing crop plants of the *Solanales* order and other orders that are devoid of toxic amounts of deleterious steroidal alkaloids typically present in edible parts of these plants. Thus, the cells and plants disclosed herein are of significant nutritional and commercial value.

[00174] Disclosed herein are an array of co-expressed genes that participate in the biosynthesis pathway of steroidal alkaloids, steroidal saponins, or triterpenoid saponins. The present invention further discloses key genes in this pathway, altering the expression of which result in concomitant alteration in the steroidal alkaloid profile, in the steroidal saponin profile, or in the triterpenoid profile. In one non-limiting example, changing the production level of steroidal alkaloids can result in an improved plant comprising elevated content of steroidal alkaloids having increased resistance to pathogens, or plants having a reduced content of these secondary compounds in the plant edible parts and thus producing improved crops. Alternatively, or additionally, in one non-limiting example, controlling the expression of genes revealed in the present invention can be used, e.g., for the production of desired steroidal alkaloids or plant-based cholesterol or other phytosterols for further use, for example, in the nutrition, cosmetic, or pharmaceutical industries. Thus, the cells and plants of the present invention are thus of significant nutritional, cosmetic, pharmaceutical, and commercial value.

[00175] Disclosed herein are genetically modified cells (e.g., plant cells, yeast cells, algal cells, bacterial cells, insect cells) and plants comprising at least one genetically modified cell, wherein the genetically modified cells are expressing at least one heterologous gene encoding an enzyme within the steroidal alkaloid/steroidal saponin/triterpenoid saponin biosynthetic pathway. These enzymes include, in some embodiments, cellulose synthase like G (CSLG) enzymes, as well as saponin beta-amyrin synthases, cytochrome P450s, glycosyltransferases, and acyltransferases.

[00176] In some embodiments, a genetically modified cell disclosed herein, expresses at least one heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase 1, and a cellulose synthase like G (CSLG). **FIGURE 1** provides a non-limiting

example of steroidal alkaloids and steroidal saponins in the triterpenoid biosynthetic pathway in Solanaceous plants (here, tomato) beginning with cholesterol, wherein a skilled artisan would recognize that the pathway may encompass production of saponin aglycones and glycosylated steroidal alkaloids biosynthetic intermediates and glycosylated saponins biosynthetic intermediates, and results in the production of a steroidal alkaloid (e.g., alpha-tomatine) or a steroidal saponin (e.g., uttroside B). **FIGURE 20A** provides a non-limiting example of a triterpenoid saponin pathway in spinach, wherein a skilled artisan would recognize that the pathway may encompass production of saponin aglycones and glycosylated saponins biosynthetic intermediates, and results in the production of a triterpenoid saponin (Yossoside V; **Compound 11**).

[00177] Altering the expression of at least one of the genes in the steroidal alkaloid/steroidal saponin/triterpenoid saponin biosynthetic pathway, in some embodiments, results in concomitant alteration in the steroidal alkaloid profile or intermediates of that pathway, in the steroidal saponin profile or intermediates of that pathway, or in the triterpenoid saponin profile or intermediates of that pathway. In some embodiments, the altered expression in a genetically modified cells comprises increased expression compared with a corresponding unmodified cell. In some embodiments, increased expression comprises de novo expression from a gene encoding an enzyme not previously present in the cell. In some embodiments, increased expression comprises an increase of expression of a gene already present within the cells. In some embodiments, increased expression results in an increased amount of the encoded enzyme.

[00178] In some embodiments, the alteration in the steroidal alkaloid profile comprising an increase in at least one steroidal alkaloid or an intermediate thereof. In some embodiments, the alteration in the steroidal alkaloid profile comprising a decrease in at least one steroidal alkaloid or an intermediate thereof.

[00179] In some embodiments, the alteration in the steroidal saponin profile comprising an increase in at least one steroidal saponin or an intermediate thereof. In some embodiments, the alteration in the steroidal saponin profile comprising a decrease in at least one steroidal saponin or an intermediate thereof.

[00180] In some embodiments, the alteration in the triterpenoid saponin profile comprising an increase in at least one triterpenoid saponin or an intermediate thereof. In some embodiments, the alteration in the triterpenoid saponin profile comprising a decrease in at least one triterpenoid saponin or an intermediate thereof.

[00181] In some embodiments, introducing a nucleotide sequence encoding at least one of the enzymes of the steroidal alkaloid/steroidal saponin/triterpenoid saponin biosynthetic pathway into a heterologous cell, (for example but not limited to a plant cell, an algal cell, a yeast cell, an insect cell,

or a bacterial cell) or an organism (for example but not limited to a plant), results in increased expression of the at least one enzyme and increased production of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin. In some embodiments, introducing a nucleotide sequence encoding at least one of the enzymes of the steroidal alkaloid/steroidal saponin/triterpenoid saponin biosynthetic pathway into a heterologous cell or organism, results in production of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin not naturally produced by the heterologous cell (for example but not limited to a plant cell, an algal cell, a yeast cell, an insect cell, or a bacterial cell) or organism (for example but not limited to a plant).

[00182] Throughout described herein are genetically modified cells and uses thereof. In some embodiments, a genetically modified cell is selected from any of an algal cell, a yeast cell, an insect cell, or a bacterial cell. One skilled in the art would know how to use the genes and nucleic acid sequences disclosed herein to genetically modified a eukaryotic cell, for example mammalian cells, or a prokaryotic cell, for example bacterium, to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. Further, one skilled in the art could use the genes and nucleic acid sequences disclosed herein to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a cell-free in vitro system. In some embodiments, the genetically modified cell is a plant cell, and said plant cell is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is a plant cell or a yeast cell and said plant cell or yeast is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is a yeast cell, and said yeast cell is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is a algal cell, and said algal cell is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is an insect cell, and said insect cell is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is a bacterium, and said bacterium is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In some embodiments, the genetically modified cell is a plant cell comprised in a plant or plant part and said plant cell is used to produce at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in said plant or plant part. In some embodiments, the genetically modified cell is a plant cell comprised in a plant or plant part and said plant cell is used to reduce the production of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in said plant or plant part. In some embodiments, the genetically modified cell is a plant cell comprised in a plant or plant part and said plant cell is used to increase the production of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in said plant or plant part.

[00183] In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from one plant species in a second plant species (see for example, **Example 23** below). In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from a plant species in a yeast cell (see for example, **Example 21** below). In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from a plant species in an algal cell. In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from a plant species in bacterial cell. In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from a plant species in an insect cell. In some embodiments, heterologous expression comprises expressing an at least one heterologous plant gene from a plant species in a cell of a plant, wherein the plant is of a different species.

[00184] In some embodiments, expressing at least one heterologous gene encoding an enzyme selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least two enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least three enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least four enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least five enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least six enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid,

steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least seven enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least eight enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least nine enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof. In some embodiments, expressing more than one heterologous gene encoding at least ten enzymes selected from a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway alters the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin produced by that pathway or alters the content of at least one biosynthetic intermediate thereof, or a combination thereof.

[00185] In some embodiments, disclosed herein is a genetically modified cell having increased expression of at least one heterologous gene compared to a corresponding unmodified cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said genetically modified cell comprises an increased content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell. In some embodiments, the genetically modified cell further expresses at least one additional heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, and an UDP-glucose 6-dehydrogenase 1, or any combination thereof.

[00186] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00187] In some embodiments, a genetically modified cell expresses a heterologous CSLG and a heterologous saponin beta-amyrin synthase. In some embodiments, a genetically modified cell

expresses a heterologous CSLG and at least one heterologous cytochrome P450. In some embodiments, a genetically modified cell expresses a heterologous CSLG and 1, 2, or 3 heterologous cytochrome P450s. In some embodiments, a genetically modified cell expresses a heterologous CSLG, a heterologous saponin beta-amyrin synthase, and 1, 2, or 3 heterologous cytochrome P450s. In some embodiments, a genetically modified cell expresses a heterologous CSLG and at least one heterologous glycosyl transferase. In some embodiments, a genetically modified cell expresses a heterologous CSLG and at least 1, 2, 3, or 4 heterologous glycosyl transferases. In some embodiments, a genetically modified cell expresses a heterologous CSLG, a heterologous saponin beta-amyrin synthase, and at least 1, 2, 3, or 4 heterologous glycosyl transferases. In some embodiments, a genetically modified cell expresses a heterologous CSLG, a heterologous saponin beta-amyrin synthase, at least 1, 2, or 3 heterologous cytochrome P450s, and at least 1, 2, 3, or 4 heterologous glycosyl transferases.

[00188] In certain instances, it may be necessary to express additional genes not specifically part of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin pathway, homologous or heterologous, in order that necessary substrates are available. In some embodiments, a genetically modified cell expresses a heterologous gene encoding an enzyme necessary for production of substrates of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway.

[00189] In some embodiments, a heterologous gene encoding an enzyme selected from a saponin beta-amyrin synthase, a cytochrome P450, a C-2hydroxylase, a C-23 oxidase, a glycosyltransferase, a UDP-glucose 6-dehydrogenase 1, an acyltransferase, or a cellulose like synthase G (CSLG) is expressed in a genetically modified cell. In some embodiments, a heterologous gene encoding an enzyme selected from a saponin beta-amyrin synthase, a cytochrome P450, C-2hydroxylase, a C-23 oxidase, a glycosyltransferase, an acyltransferase, or a cellulose like synthase G (CSLG) is expressed in a genetically modified cell. In some embodiments, a heterologous gene encoding an enzyme selected from a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, a UDP-glycosyltransferase, a fucosyltransferase, a xylosyltransferase, an acyltransferase, a BAHD-acyltransferase, or a cellulose like synthase G (CSLG) in expressed in a genetically modified cell.

[00190] In some embodiments, a genetically modified cell expresses at least one heterologous gene encoding an enzyme comprising a CSLG activity. In some embodiments, a genetically modified cell expresses at least one heterologous gene encoding an enzyme comprising a saponin beta-amyrin synthase activity. In some embodiments, a genetically modified cell expresses at least one heterologous gene encoding an enzyme comprising a cytochrome P450 activity. In some embodiments, a genetically modified cell expresses at least two heterologous gene encoding two enzymes comprising cytochrome P450 activity. In some embodiments, a genetically modified cell

expresses at least three heterologous gene encoding three enzymes comprising cytochrome P450 activity. In some embodiments, an enzyme comprising cytochrome p450 activity comprises a C-2 hydroxylase activity. In some embodiments, an enzyme comprising cytochrome p450 activity comprises a C-23 oxidase activity.

[00191] In some embodiments, a genetically modified cell expresses heterologous genes encoding enzymes comprising a saponin beta-amyrin synthase activity, cytochrome P450 activities, and CSLG activity, wherein the cytochrome P450 enzymes may comprise an enzyme providing a C-2 hydroxylase activity, an enzyme providing a C-23 oxidase activity, a combination of enzymes providing cytochrome P450 activity, C-2 hydroxylase activity, and C-23 oxidase activity.

[00192] In some embodiments, a genetically modified cell expresses at least one heterologous gene encoding a glycosyltransferase enzyme activity. In some embodiments, the glycosyltransferase enzyme comprises a UDP-glycosyltransferase, a fucosyltransferase, or a xylosyltransferase. In some embodiments, a genetically modified cell expresses multiple heterologous genes encoding glycosyltransferase enzymes. In some embodiments, a genetically modified cell expresses multiple heterologous genes encoding glycosyltransferases, selected from a UDP-glycosyltransferase, a fucosyltransferase, or a xylosyltransferase, or a combination thereof. In some embodiments, a genetically modified cell expresses multiple heterologous genes encoding glycosyltransferases, selected from a combination of known glycosyltransferases known in the art. In some embodiments, a genetically modified cell expresses a heterologous gene encoding a CSLG activity and a heterologous gene encoding at least one glycosyltransferase enzyme activity. In some embodiments, a genetically modified cell expresses a heterologous gene encoding a CSLG activity and a heterologous gene encoding at least one glycosyltransferase enzyme activity selected from a UDP-glycosyltransferase, a fucosyltransferase, or a xylosyltransferase, or a combination thereof. In some embodiments, a genetically modified cell expresses a heterologous gene encoding a CSLG activity and a heterologous gene encoding at least one glycosyltransferase enzyme activity selected from a UDP-glycosyltransferase, a fucosyltransferase, or a xylosyltransferase, or any glycosyltransferase known in the art, or a combination thereof.

[00193] In some embodiments, a genetically modified cell expresses a heterologous gene encoding an acyltransferase. In some embodiments, a genetically modified cell expresses a heterologous gene encoding a benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamoyl/benzoyl-, deacetylvindoline acetyltransferase (BAHD) acyltransferase. In some embodiments, a genetically modified cell expresses a heterologous gene encoding an enzyme providing a CSLG activity and a heterologous gene encoding an enzyme encoding an acetyltransferase activity. In some embodiments, a genetically modified cell expresses a heterologous

gene encoding an enzyme providing a CSLG activity and a heterologous gene encoding an enzyme encoding a BAHD acetyltransferase activity.

[00194] In some embodiments, a genetically modified cell expresses a heterologous gene encoding the encodes an enzyme activity for a triterpenoid saponin substrate, wherein the enzyme comprises a glycosyltransferase. In some embodiments, a genetically modified cell expresses a heterologous gene encoding the encodes an enzyme activity for a triterpenoid saponin substrate, wherein the enzyme comprises a UDP-glucose 6-dehydrogenase 1. In some embodiments, for example when the cell is a bacterial cell, a heterologous gene encodes a Squalene epoxidase, or a Cytochrome P450 reductase, or a UDP-glucose 6-dehydrogenase 1, or any combination thereof.

[00195] In some embodiments, disclosed herein is a genetically modified plant comprising an at least one cell having altered expression of at least a CSLG gene compared to the expression of CSLG in a corresponding unmodified cell, wherein the at least one cell comprises a heterologous CSLG gene, and said altered expression comprises a de novo expression of said heterologous gene. In some embodiments, disclosed herein is a genetically modified plant comprising an at least one cell having altered expression of at least a CSLG gene compared to the expression of CSLG in a corresponding unmodified cell, wherein the at least one cell comprises a heterologous CSLG gene, and said altered expression comprises expression of a heterologous gene in the presence of an endogenous gene, which may be functional or non-functional.

[00196] In some embodiments, the expression of a heterologous CSLG gene is a genetically modified plant comprises increased expression compared to a corresponding unmodified cell, and wherein said cell in said plant comprises altered content of at least a triterpenoid saponin, metabolite thereof, derivative thereof, or biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00197] In some embodiments, disclosed herein is an array of enzymes that participate in the biosynthetic pathway of triterpenoid saponins, wherein the biosynthesis pathway includes production of saponin aglycones and glycosylated saponins, and glycosylated triterpenoid saponins. In some embodiments, disclosed herein is an array of genes encoding that enzymes that participate in the biosynthetic pathway of triterpenoid saponins, wherein the biosynthesis pathway includes production of saponin aglycones and glycosylated saponins, and glycosylated triterpenoid saponins, including but not limited to 3-O-[β-D-xylopyranosyl-(1–>3)-β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-4-acetyl-β-D-fucopyranosyl]-medicagenic acid (Yossoside V; **Compound 11**); 3-O-[β-D-xylopyranosyl-(1–>3)-β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid (Yossoside IV; **Compound 10**); 3-O-[β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–

>4)-α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid (Yossoside III; **Compound 9**); 3-O-[β-D-glucuronopyranosyl]-28-O-[α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid (Yossoside II; **Compound 8**); and 3-O-β-D-glucuronopyranosyl-28-O-β-D-fucopyranosyl-medicagenic acid (Yossoside I; **Compound 7**) (see **FIGURE 20A**).

[00198] For example, but not limited to the enzymes active in the spinach triterpenoid biosynthetic pathway of **Compound 11**, which are described in detail in the Examples. **Table 16** provided in **Example 20**, provides the amino acid and nucleic acid sequences of the enzymes in the spinach biosynthetic pathway, wherein CSLG enzyme adds a glucuronic acid to medicagenic acid to make the triterpenoid saponin medicagenic acid 3-glucuronide. **Table 19** disclosed in **Example 24** provides the amino acid sequence of a CSLG isolated from Arabidopsis. **Table 17** disclosed in the Examples provides the nucleic acid and amino acid sequences of CSLG homologs from other plant species.

[00199] In some embodiments, a genetically modified cell disclosed herein expresses at least one heterologous gene encoding an enzyme comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses an array of heterologous gene encoding enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least two heterologous gene encoding two enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least three heterologous gene encoding three enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least four heterologous gene encoding four enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least 5 heterologous gene encoding five enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least six heterologous gene encoding six enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least seven heterologous gene encoding seven enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least eight heterologous gene encoding eight enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell disclosed herein expresses at least nine heterologous gene encoding nine enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell

disclosed herein expresses at least ten heterologous gene encoding ten enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway.

[00200] In some embodiments, disclosed herein is a genetically modified plant comprising at least one cell having altered expression of at least a cellulose synthase like G (CSLG) gene compared to the expression of CSLG in a corresponding unmodified plant, and wherein the genetically modified plant has an altered content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant. In some embodiments, the CSLG gene comprises the endogenous CSLG gene. In some embodiments, the CSLG gene comprises a heterologous CSLG gene.

[00201] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00202] In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least one heterologous gene encoding an enzyme comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses an array of heterologous gene encoding enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least two heterologous gene encoding two enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least three heterologous gene encoding three enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least four heterologous gene encoding four enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least 5 heterologous gene encoding five enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least six heterologous gene encoding six enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least seven heterologous gene encoding seven enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed

herein expresses at least eight heterologous gene encoding eight enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least nine heterologous gene encoding nine enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway. In some embodiments, a genetically modified plant or plant part disclosed herein expresses at least ten heterologous gene encoding ten enzymes comprised in a steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway.

[00203] In some embodiments, heterologous genes encoding enzymes of the steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway comprising enzymes selected from a beta-amyrin synthase, a cytochrome P450, multiple cytochrome P450s, a glycosyltransferase, multiple glycosyltransferases, an acyltransferase, or a CSLG, or any combination thereof. In some embodiments, heterologous genes encoding enzymes of the steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway comprising enzyme selected from a beta-amyrin synthase, a cytochrome P450, multiple cytochrome P450s, a glycosyltransferase, multiple glycosyltransferases, an acyltransferase, or a CSLG, or any combination thereof, and in addition express at least one heterologous gene necessary to provide a biosynthetic pathway substrate or substrates. In some embodiments, heterologous genes encode enzymes of the steroidal alkaloid, steroidal saponin, or triterpenoid biosynthetic pathway comprising a beta-amyrin synthase, a cytochrome P450, multiple cytochrome P450s, a glycosyltransferase, multiple glycosyltransferases, an acyltransferase, or a CSLG, or any combination thereof, and in addition express a heterologous gene necessary to provide biosynthetic pathway substrates, for example but not limited to a UDP-glucose 6-dehydrogenase 1 enzyme.

[00204] In some embodiments, a genetically modified cell expresses at least one heterologous gene encoding an enzyme of the steroidal alkaloid/steroidal saponin/triterpenoid saponin biosynthetic pathway, said at least one heterologous gene encoding a CSLG enzyme.

[00205] In some embodiments, the at least one heterologous gene encodes a cytochrome P450 wherein said cytochrome P450 comprises a C2-hydroxylase or a C-23 oxidase, or a glycosyltransferase wherein said glycosyltransferase comprises a fructosyltransferase, a xylosyltransferase, or a UDP-glycosyltransferase, or a acyltransferase wherein said acyltransferase comprises a benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamyol/benzoyl-, deacetylvindoline (BAHD) acetyletransferase, or any combination thereof.

[00206] In some embodiments, expression of a heterologous gene or genes comprises increased expression compared to a corresponding unmodified cell. In some embodiments, increased expression of a gene results in increased content of an enzyme encoded by the gene, compared to the

content of the enzyme in a corresponding unmodified cell. One skilled in the art would appreciate that increased expression may, in some embodiments, comprise overexpression of the gene or genes, with resultant increased presence of an enzyme of enzymes encoded by the gene or genes, compared with a corresponding unmodified cell. In some embodiments, increased expression may comprise expression of a gene or genes, and the production of a resultant enzyme activity or activities, not naturally found in the cell or organism, for example but not limited to a plant cell, a plant, a yeast cell, an algal cell, an insect cell, or a bacterial cell.

[00207] A skilled artisan would appreciate that a corresponding unmodified cell may in certain embodiments provide a control cell, wherein the heterologous gene is not expressed or is not overexpressed.

[00208] In some embodiments, expression of a gene or genes in a heterologous cell comprises transient expression of the gene or genes. In some embodiments, expression of a gene or genes in a heterologous cell comprises constitutive expression of the gene or genes. In some embodiments, expression of a gene or genes in a heterologous cell comprises transient expression, or constitutive expression, or a combination thereof, of the gene or genes.

[00209] A skilled artisan would appreciate that transient expression encompasses the temporary expression of a gene or genes that are expressed for a short time after a nucleic acid has been introduced into eukaryotic cells, for example a plant cell or a yeast cell. Conversely, constitutive expression, also known as stable expression, encompasses continuous expression of a gene or genes.

[00210] One of ordinary skill in the art would appreciate that the term “gene” may encompass a nucleic acid (e.g., DNA or RNA) sequence that comprises coding sequences necessary for the production of RNA or a polypeptide. A polypeptide can be encoded by a full-length coding sequence or by any part thereof. The term "parts thereof" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides to the entire gene sequence minus one nucleotide. Thus, "a nucleic acid sequence comprising at least a part of a gene" may comprise fragments of the gene or the entire gene.

[00211] The skilled artisan would appreciate that the term "gene" optionally also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding regions and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences.

[00212] In some embodiments, a gene comprises DNA sequence comprising the coding region. In

some embodiments, a gene comprises DNA sequence comprising upstream and downstream regions, as well as the coding region, which comprises exons and any intervening introns of the gene. In some embodiments, upstream and downstream regions comprise non-coding regulatory regions. In some embodiments, upstream and downstream regions comprise regulatory sequences, for example but not limited to promoters, enhancers, and silencers. Non-limiting examples of regulatory sequences include, but are not limited to, AGGA box, TATA box, Inr, DPE, ZmUbi1, PvUbi1, PvUbi2, CaMV, 35S, OsAct1, zE19, E8, TA29, A9, pDJ3S, B33, PAT1, alcA, G-box, ABRE, DRE, and PCNA.

[00213] Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within a cell, for example but not limited to a yeast, an algal, an insect, a bacterium, or a plant cell or a plant described herein. Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within a plant cell. Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within an algal cell. Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within a yeast cell. Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within an insect cell. Regulatory regions, may in some embodiments, increase or decrease the expression of specific genes within a bacterium.

[00214] In some embodiments, a gene comprises the coding regions of the gene, which comprises exons and any intervening introns of the gene. In some embodiments, a gene comprises its regulatory sequences. In some embodiments, a gene comprises the gene promoter. In some embodiments, a gene comprises its enhancer regions. In some embodiments, a gene comprises 5′ non-coding sequences. In some embodiments, a gene comprises 3′ non-coding sequences.

[00215] In some embodiments, the skilled artisan would appreciate that DNA comprises a gene, which may include upstream and downstream sequences, as well as the coding region of the gene. In other embodiments, DNA comprises a cDNA (complementary DNA). One of ordinary skill in the art would appreciate that cDNA may encompass synthetic DNA reverse transcribed from RNA through the action of a reverse transcriptase. The cDNA may be single stranded or double stranded and can include strands that have either or both of a sequence that is substantially identical to a part of the RNA sequence or a complement to a part of the RNA sequence. Further, cDNA may include upstream and downstream regulatory sequences. In still other embodiments, DNA comprises a CDS (complete coding sequence). One of ordinary skill in the art would appreciate that CDS may encompass a DNA sequence, which encodes a full-length protein or polypeptide. A CDS typically begins with a start codon ("ATG") and ends at (or one before) the first in-frame stop codon ("TAA", "TAG", or "TGA"). The skilled artisan would recognize that a cDNA, in one embodiment, comprises a CDS.

[00216] The terms "polynucleotide", "polynucleotide sequence", "nucleic acid sequence", and

"isolated polynucleotide" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA or hybrid thereof, that is single- or double-stranded, linear or branched, and that optionally contains synthetic, non-natural or altered nucleotide bases. The terms also encompass RNA/DNA hybrids.

[00217] The terms “complementary” or “complement thereof” are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

[00218] The term "construct" as used herein refers to an artificially assembled or isolated nucleic acid molecule which includes the polynucleotide of interest. In general, a construct may include the polynucleotide or polynucleotides of interest, a marker gene, which in some cases can also be a gene of interest and appropriate regulatory sequences. It should be appreciated that the inclusion of regulatory sequences in a construct is optional, for example, such sequences may not be required in situations where the regulatory sequences of a host cell are to be used. Regulatory elements include, but are not limited to, a promoter, an enhance, an origin of replication, a transcription termination sequence, a polyadenylation signal, and the like. The term “construct” includes vectors but should not be seen as being limited thereto.

[00219] The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is regulated by the other. For example, a promoter is operably linked with a coding sequence when it is capable of regulating the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation. In some embodiments, the terms “operably linked” and “functionally linked” may be used interchangeably having all the same meanings and qualities.

[00220] The terms "promoter element," "promoter," or "promoter sequence" as used herein, refer to a DNA sequence that is generally located at the 5' end (i.e. precedes) the coding region of a DNA polymer. The location of most promoters known in nature precedes the transcribed region. The promoter functions as a switch, activating the expression of a gene. If the gene is activated, it is said to be transcribed, or participating in transcription. Transcription involves the synthesis of mRNA from the gene. The promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA.

[00221] A polynucleotide sequence comprising a heterologous gene may be expressed in, for example, a plant cell, an algal cell, a yeast cell, an insect cell, or a bacterium under the control of a

promoter that directs constitutive expression or regulated (transient) expression. In some embodiments, the plant cell is part of a whole plant or a portion thereof. Regulated expression comprises temporally or spatially regulated expression and any other form of inducible or repressible expression. Temporally means that the expression is induced at a certain time point, for instance, when a certain growth rate of the plant cell culture is obtained (e.g., the promoter is induced only in the stationary phase or at a certain stage of development). Spatially means that the promoter is only active in specific organs, tissues, or cells (e.g., for plants only in roots, leaves, epidermis, guard cells or the like). Other examples of regulated expression comprise promoters whose activity is induced or repressed by adding chemical or physical stimuli to the plant cell or yeast cell or an algal cell or an insect cell or a bacterium. In some embodiments, a promoter is an inducible promoter. In some embodiments, a promoter is a constitutive promoter. In some embodiments regulated expression of a heterologous gene of the triterpenoid biosynthetic pathway, for example but not limited to CSLG, comprises inducing or repressing gene expression by adding chemical or physical stimuli to a plant cell. In some embodiments regulated expression of a heterologous gene of the triterpenoid biosynthetic pathway, for example but not limited to CSLG, comprises inducing or repressing gene expression by adding chemical or physical stimuli to a yeast cell. In some embodiments regulated expression of a heterologous gene of the triterpenoid biosynthetic pathway, for example but not limited to CSLG, comprises inducing or repressing gene expression by adding chemical or physical stimuli to an algal cell. In some embodiments regulated expression of a heterologous gene of the triterpenoid biosynthetic pathway, for example but not limited to CSLG, comprises inducing or repressing gene expression by adding chemical or physical stimuli to an insect cell. In some embodiments regulated expression of a heterologous gene of the triterpenoid biosynthetic pathway, for example but not limited to CSLG, comprises inducing or repressing gene expression by adding chemical or physical stimuli to a bacterium.

[00222] In some embodiments, the expression is under control of environmental, hormonal, chemical, and/or developmental signals. Such promoters for plant cells include promoters that are regulated by (1) heat, (2) light, (3) hormones, such as abscisic acid, and methyl jasmonate (4) wounding or (5) chemicals such as salicylic acid, chitosans or metals. A constitutive promoter directs expression in a wide range of cells under a wide range of conditions. Examples of constitutive plant promoters useful for expressing heterologous polypeptides in plant cells include, but are not limited to, the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues including monocots, the nopaline synthase promoter and the octopine synthase promoter. In some embodiments, a promoter is an inducible promoter. In some embodiments provided herein, inducible plant promoters, include but are not limited to GAL1,

GAL10, PHO5 and CUP1. In some embodiments, a yeast promoter is constitutive promoter. In some embodiments provided herein, constitutive yeast promoters, include but are not limited to PGK1, TDHS, ADH1, CYC1, ACT1, TEF1. In some embodiments, an algal promoter is an inducible promoter. In some embodiments, provided herein algal promoters may be regulated by changes of light, pH, salinity, temperature, or nutrients. In some embodiments, an inducible algal promoter responds to nitrogen levels, nitrogen stress, or nitrogen starvation. In some embodiments, an algal promoter is a constitutive promoter. In some embodiments, an inducible promoter for use in insect cells responds to environmental or nutritional signals.

[00223] An expression cassette is usually provided in a DNA or RNA construct comprising the at least one heterologous gene that is typically called an "expression vector," which is any genetic element, e.g., a plasmid, a chromosome, a virus, behaving either as an autonomous unit of polynucleotide replication within a cell (i.e., capable of replication under its own control) or being rendered capable of replication by insertion into a host cell chromosome, having attached to it another polynucleotide segment, so as to bring about the replication and/or expression of the attached segment. Suitable vectors include, but are not limited to, plasmids, bacteriophages, cosmids, plant viruses, and artificial chromosomes. The expression cassette may be provided in a DNA construct, which also has at least one replication system. In addition to the replication system, there will frequently be at least one marker present, which may be useful in one or more hosts, or different markers for individual hosts.

[00224] In another embodiment, selection genes provide selective markers that confer resistance to a selective agent, such as an antibiotic or herbicide. Potentially transformed cells are exposed to the selective agent. In the population of surviving cells will be those cells where, generally, the resistance-conferring gene has been integrated and expressed at sufficient levels to permit cell survival. Cells may be tested further to confirm stable integration of the exogenous DNA.

[00225] Useful selective marker genes include those conferring resistance to antibiotics such as kanamycin (nptII), hygromycin B (aph IV), nptII, hpt, aadA and gentamycin (aac3 and aacC4) or resistance to herbicides such as glufosinate (bar or pat) and glyphosate (EPSPS). In another embodiment, the selection gene is an antimetabolite. In one embodiment, the antimetabolite is dhf.

[00226] Screenable markers which provide an ability to visually identify transformants can also be employed, e.g., a gene expressing a colored or fluorescent protein such as a cat, lacZ, uidA, luciferase or green fluorescent protein (GFP) or a gene expressing a beta-glucuronidase or uidA gene (GUS) for which various chromogenic substrates are known. It is also contemplated that combinations of screenable and selectable markers will be useful for identification of transformed cells.

[00227] In one embodiment, the selection gene is a positive selectable marker gene that is

conditional on non-toxic agents that may be substrates for growth or that induce growth and differentiation of the transformed tissues.

[00228] In one embodiment, plant cells that survive exposure to the selective agent, or cells that have been scored positive in a screening assay, may be cultured in regeneration media and allowed to mature into plants. Developing plantlets can be transferred to soil less plant growth mix, and hardened off, e.g., in an environmentally controlled chamber at about 85% relative humidity, 600 ppm CO2, and 25-250 microeinsteins m−2 s−1 of light, prior to transfer to a greenhouse or growth chamber for maturation. Plants are preferably matured either in a growth chamber or greenhouse. Plants are regenerated from about 6 weeks to 10 months after a transformant is identified, depending on the initial tissue. During regeneration, cells are grown to plants on solid media at about 19 to 28° C. After regenerating plants have reached the stage of shoot and root development, they may be transferred to a greenhouse for further growth and testing. Plants may be pollinated using conventional plant breeding methods known to those of skill in the art and seed produced.

[00229] Progeny may be recovered from transformed plants and tested for expression of the exogenous recombinant polynucleotide. Useful assays include, for example, "molecular biological" assays, such as Southern and Northern blotting and PCR; "biochemical" assays, such as detecting the presence of RNA, e.g. double stranded RNA, or a protein product, e.g., by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and also, by analyzing the phenotype of the whole regenerated plant.

[00230] One of skill in the art will be able to select an appropriate vector for introducing the encoding nucleic acid sequence in a relatively intact state. Thus, any vector which will produce a host cell, e.g., plant protoplast, carrying the introduced encoding nucleic acid should be sufficient. The selection of the vector, or whether to use a vector, is typically guided by the method of transformation selected.

[00231] For plant cells, markers may a) code for protection against a biocide, such as antibiotics, toxins, heavy metals, certain sugars or the like; b) provide complementation, by imparting prototrophy to an auxotrophic host; or c) provide a visible phenotype through the production of a novel compound in the plant. In some embodiments, genes that may be employed include neomycin phosphotransferase (NPTII), hygromycin phosphotransferase (HPT), chloramphenicol acetyltransferase (CAT), nitrilase, and the gentamicin resistance gene. For plant host selection, non-limiting examples of suitable markers are β-glucuronidase, providing indigo production, luciferase, providing visible light production, Green Fluorescent Protein and variants thereof, NPTII, providing kanamycin resistance or G418 resistance, HPT, providing hygromycin resistance, and the mutated aroA gene, providing glyphosate resistance.

[00232] For yeast cells, markers may comprise marker providing complementation, by imparting prototrophy to an auxotrophic host, for example but not limited to HIS, LEU, TRP, and URA marker genes encoding the necessary complementary activity. In some embodiments, yeast markers provide resistance to a compound toxic to the yeast, for example but not limited to kanMX4 – resistance to G418; natMX4 – resistance to nourseothricin; hphMX4 – resistance to hygromycin B; patMX3 – resistance to glufosinate; and ZEO – resistance to zeocin.

[00233] Promoter activity encompasses the extent of transcription of a polynucleotide sequence, homologue, variant or fragment thereof that is operably linked to the promoter whose promoter activity is being measured. The promoter activity may be measured directly by measuring the amount of RNA transcript produced, for example, by Northern blot or indirectly by measuring the product coded for by the RNA transcript, such as when a reporter gene is linked to the promoter.

[00234] In some embodiments, an at least one heterologous gene encoding an enzyme, disclosed herein, is operably linked to a promoter. In some embodiments, operable linkage to the promoter results in transient expression. In some embodiments, operable linkage to the promoter results in constitutive expression. In some embodiments, one heterologous gene may be operably linked to a constitutive promoter while another heterologous gene is operably linked to a promoter providing transient expression. In some embodiments, a polynucleotide sequence comprises the at least one heterologous gene encoding an enzyme, wherein said polynucleotide is optionally comprised within a vector.

[00235] As used herein, the term an "enhancer" refers to a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter.

[00236] The term "expression", as used herein, refers to the production of a functional end-product e.g., an mRNA or a protein. In some embodiments, a genetically modified cell expresses at least one heterologous gene, wherein the functional end product comprises an mRNA, which in turn is transcribed to produce a functional end product comprising a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin.

[00237] The term “genetically modified cell” refers to a cell genetically modified by man. In some embodiments, the genetic modification includes transforming a plant cell or a yeast cell or an algal cell or an insect cell or a bactierium with heterologous polynucleotide comprising a gene encoding an enzyme. In some embodiments, the genetic modification includes transforming a plant cell or a yeast cell with heterologous polynucleotide comprising a gene encoding an enzyme active in a triterpenoid saponin biosynthetic pathway. A “genetically modified cell” and a “corresponding unmodified cell” as used herein refer to a genetically modified cell and to a cell of the same type lacking said

modification, respectively.

[00238] In some embodiments, the "genetically modified cell" and the "corresponding unmodified cell" are, respectively, a genetically modified plant cell and a corresponding unmodified plant cell of the same type. In other embodiments, the "genetically modified cell" and the "corresponding unmodified cell" are, respectively, a genetically modified yeast cell and a corresponding unmodified yeast cell of the same type.

[00239] The term "genetically modified plant" refers to a plant comprising at least one cell genetically modified by man. The genetic modification includes modification of an endogenous gene(s), for example by introducing mutation(s) deletions, insertions, transposable element(s) and the like into an endogenous polynucleotide or gene of interest. Additionally, or alternatively, the genetic modification includes transforming the plant cell with heterologous polynucleotide. A "genetically modified plant" and a "corresponding unmodified plant" as used herein refer to a plant comprising at least one genetically modified cell and to a plant of the same type lacking said modification, respectively.

[00240] In some embodiments, suspensions of genetically modified cells and tissue cultures derived from the genetically modified cells are also encompassed. The cell suspension or tissue culture can be used for the production of desired steroidal alkaloids metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; steroidal saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; or triterpenoid saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof. In some embodiments, the genetically modified cell or tissue culture is used for regenerating a genetically modified plant having at least one cell having altered expression of a CSLG enzyme.

[00241] One of ordinary skill in the art would appreciate that some embodiments further encompass seeds of the genetically modified plant, wherein plants grown from said seeds have altered expression of at least one CSLG in at least one cell compared to plants grown from corresponding unmodified seeds or seeds from a corresponding unmodified plant, thereby having at least one cell having altered content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin.

[00242] If transformation techniques require use of tissue culture, transformed cells may be regenerated into plants in accordance with techniques well known to those of skill in the art. The regenerated plants may then be grown and crossed with the same or different plant varieties using traditional breeding techniques to produce seeds, which are then selected under the appropriate conditions.

[00243] One of ordinary skill in the art would appreciate that a genetically modified plant may

encompass a plant comprising at least one cell genetically modified by man. In some embodiment, the genetic modification includes transforming at least one plant cell or yeast cell or algal cell or insect cell or bacterium with a heterologous polynucleotide comprising a heterologous gene described herein. In some embodiment, the genetic modification includes transforming at least one plant cell or yeast cell or algal cell or insect cell or bacterium with a heterologous polynucleotide comprising more than one heterologous gene, each encoding an enzyme. In some embodiment, the genetic modification includes transforming at least one plant cell or yeast cell or algal cell or insect cell or bacterium with multiple heterologous polynucleotides encoding multiple heterologous genes encoding enzymes.

[00244] The skilled artisan would appreciate that a genetically modified plant comprising transforming at least one plant cell with a heterologous polynucleotide or multiple heterologous polynucleotides may in certain embodiments be termed a “transgenic plant”.

[00245] A skilled artisan would appreciate that a comparison of a “genetically modified plant” to a “corresponding unmodified plant” as used herein encompasses comparing a plant comprising at least one genetically modified cell and to a plant of the same type lacking the modification.

[00246] The skilled artisan would appreciate that the term "transgenic" when used in reference to a plant as disclosed herein encompasses a plant that contains at least one heterologous transcribable polynucleotide in one or more of its cells. The skilled artisan would appreciate that the term "transgenic" when used in reference to a yeast as disclosed herein encompasses a yeast that contains at least one heterologous transcribable polynucleotide. The term "transgenic material" encompasses broadly a plant or a part thereof, including at least one cell, multiple cells or tissues that contain at least one heterologous polynucleotide in at least one of cell. Thus, comparison of a “transgenic plant” and a “corresponding non transgenic plant”, or of a “genetically modified plant comprising at least one cell having altered expression, wherein said plant comprising at least one cell comprising a heterologous transcribable polynucleotide” and a “corresponding un modified plant” encompasses comparison of the “transgenic plant” or “genetically modified plant” to a plant of the same type lacking said heterologous transcribable polynucleotide. A skilled artisan would appreciate that, in some embodiments, a “transcribable polynucleotide” comprises a polynucleotide that can be transcribed into an RNA molecule by an RNA polymerase.

[00247] The terms "transformants" or "transformed cells" include the primary transformed cell and cultures derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

[00248] Transformation of a cell may be stable or transient. The term "transient transformation" or

"transiently transformed" refers to the introduction of one or more exogenous polynucleotides into a cell in the absence of integration of the exogenous polynucleotide into the host cell's genome. In contrast, the term "stable transformation" or "stably transformed" refers to the introduction and integration of one or more exogenous polynucleotides into the genome of a cell. The term "stable transformant" refers to a cell which has stably integrated one or more exogenous polynucleotides into the genomic or organellar DNA. It is to be understood that an organism or its cell transformed with the nucleic acids, constructs and/or vectors disclosed herein can be transiently as well as stably transformed.

[00249] The skilled artisan would appreciate that the term "construct" may encompass an artificially assembled or isolated nucleic acid molecule which includes the polynucleotide of interest. In general, a construct may include the polynucleotide or polynucleotides of interest, a marker gene which in some cases can also be a gene of interest and appropriate regulatory sequences. It should be appreciated that the inclusion of regulatory sequences in a construct is optional, for example, such sequences may not be required in situations where the regulatory sequences of a host cell are to be used. The term construct includes vectors but should not be seen as being limited thereto.

[00250] Additionally, or alternatively, in some embodiments, the genetic modification includes transforming at least one plant cell with a heterologous polynucleotide or multiple heterologous polynucleotides. The skilled artisan would appreciate that a genetically modified plant comprising transforming at least one plant cell with a heterologous polynucleotide or multiple heterologous polynucleotides may in certain embodiments be termed a "transgenic plant".

[00251] Viral vectors are useful for transformation of more transformation-resistant plants (e.g., soybean or common bean). In some embodiments, viral vectors, such as bean pod mottle virus (BPMV; genus Comovirus) vectors, are used for foreign gene expression and virus-induced gene silencing (VIGS) (Zhang et al. (May 2010) Plant Physiol. 153: 52-65 ["Zhang 2010"])). Cells are transformed, e.g., via biolistics or via direct DNA-rubbing inoculation (Zhang 2010).

[00252] In one embodiment, a gene gun or a biolistic particle delivery system (biolistics) is used for plant transformation to deliver exogenous DNA (transgenes) to cells (Rech et al. (2008) Nature Protocols 3(3): 410-418 ["Rech 2008"]). In some embodiments, the plasmid is designed and apical meristems of plants (e.g., soybean, bean, cotton) are bombarded with microparticle-coated DNA, followed by in vitro culture and selection of transgenic plants (Rech 2008). In other embodiments, a callus of undifferentiated plant cells or a group of immature embryos growing on gel medium in vitro. In some embodiments, the cells are then treated with a series of plant hormones, such as auxins or gibberellins to obtain plants. Steroidal alkaloids metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; steroidal saponins, metabolites thereof,

derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; or triterpenoid saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof, can be produced through such transgenic plants or plant cells.

[00253] "Transient expression" of the proteins may be achieved by various means known in the art. In one embodiment, transient expression of the proteins is achieved by the use of genetically modified viruses. In some embodiments, agroinfiltration is used to induce transient expression of genes in a plant or an isolated leaf or another portion of a plant. A suspension of *Agrobacterium* (e.g., *Agrobacterium tumefaciens*) is introduced into the plant by, e.g., direct injection or vacuum filtration, or is brought into association with plant cells immobilized on a porous support (plant cell packs). The bacteria transfer the desired gene into the plant cells via transfer of Ti plasmid-derived T-DNA.

[00254] In some embodiments, the method of transformation of algae comprises any of the methods as described hereinabove. In one embodiment, transformation of algae is accomplished using glass bead-assisted transformation, particle gun-mediated (biolistic) transformation, treatment with cellulolytic enzymes to weaken their cell walls, or homologous recombination.

[00255] In another aspect, the nucleic acids disclosed herein can be transformed into algae. In one embodiment, the alga is a single cell alga. In another embodiment, the alga is a multi-cellular alga. In one embodiment, the alga is a cyanobacterium, diatom, *Chlamydomonas*, *Dunaliella*, or *Hematococus*. The genes can be over expressed in algae. Steroidal alkaloids metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; steroidal saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; or triterpenoid saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof, can be produced through such transgenic algae. Method for algae transformation are well known in the art and fully described in U.S. Patent Application Publications US 20150011008; US 20150004704; US 20130130389; US 20120094385; US 20120034698; US 20110300633; and US 20040133937, which are incorporated by reference herein in their entirety. Further, cultivation of microalgae is well known in the art and has been described in Vuppaladadiyam et al., (2018) "Microalgae cultivation and metabolites production: a comprehensive review." Biofuls, Bioprod. Bioref. 12:304-324, which is incorporated by reference herein in its entirety.

[00256] In another aspect, the nucleic acids disclosed herein can be transformed into yeast. The genes can be over expressed in yeast. Methods for yeast transformation are described herein below and are well known in the art and fully described in U.S. Patent Application Publications US 20090264320; US 20010031724; US 20030049785; US 20050158861; US 20070264716; US 20090325247; and US 20100190223, which are incorporated by reference herein in their entirety.

Steroidal alkaloids metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; steroidal saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; or triterpenoid saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof, can be produced through such transgenic yeast.

[00257] In another embodiment, the nucleic acids disclosed herein can be transformed into a virus. In another embodiment, the nucleic acids may be over-expressed in a virus. Steroidal alkaloids metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; steroidal saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof; or triterpenoid saponins, metabolites thereof, derivatives thereof, or biosynthetic intermediates thereof, or a combination thereof, can be produced through such viral over expression.

[00258] In some embodiments, in bacterial systems, a number of expression vectors can be advantageously selected depending upon the use intended for the polypeptide expressed. In one embodiment, large quantities of polypeptide are desired. In one embodiment, vectors that direct the expression of high levels of the protein product, possibly as a fusion with a hydrophobic signal sequence, which directs the expressed product into the periplasm of the bacteria or the culture medium where the protein product is readily purified are desired. In one embodiment, certain fusion protein engineered with a specific cleavage site to aid in recovery of the polypeptide. In one embodiment, vectors adaptable to such manipulation include, but are not limited to, the pET series of *Escherichia coli* expression vectors (Studier et al., Methods in Enzymol. 185:60-89 1990).

[00259] In some embodiments, a polynucleotide sequence comprises the at least one heterologous gene. In some embodiments, the polynucleotide sequence encodes an enzyme. In some embodiments, the polynucleotide sequence encodes an enzyme of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway. In some embodiments, the polynucleotide sequence encodes an enzyme necessary for the production of a substrate or substrates of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway.

[00260] In some embodiments, a polynucleotide sequence comprises the at least 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more heterologous genes. In some embodiments, the polynucleotide sequence encodes 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more enzyme, wherein each gene encodes a single enzyme. In some embodiments, the polynucleotide sequence encodes 2, 3, 4, 5, 6, 7, 8, 9, or 10 enzymes of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway. In some embodiments, the polynucleotide sequence encodes 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more enzymes of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin pathway and an additional enzyme necessary for the

production of a substrate of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway. In some embodiments, the polynucleotide sequence encodes 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more enzymes of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin pathway and additional enzymes necessary for the production of substrates of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway.

[00261] In some embodiments, the term "a nucleotide sequence" or "a polynucleotide sequence" encompasses a single contiguous polynucleotide sequence, and in other embodiments, the term encompasses multiple contiguous nucleotide sequences.

[00262] Based on the co-expressed gene array disclosed herein, a pathway from cholesterol to, e.g., α-tomatine is proposed (**FIGURE 1**). It has been previously described that cholesterol is hydroxylated at C22 by GAME7 (US 2012/0159676) followed by GAME8 hydroxylation at the C26 position. The 22,26-dihydroxycholesterol is than hydroxylated at C16 and oxidized at C22 followed by closure of the E-ring by GAME11 and GAME6 to form the furostanol-type aglycone. This order of reactions is supported by the finding herein showing the accumulation of cholestanol-type saponins, lacking hydroxylation at C16 and the hemi-acetal E-ring when silencing GAME11 (**FIGURES 8A-D**). The furostanol-intermediate is oxidized by GAME4 to its 26-aldehyde which is the substrate for transamination catalyzed by GAME12. Nucleophilic attack of the amino-nitrogen at C22 leads to the formation of tomatidenol which is dehydrogenated to tomatidine. Tomatidine is subsequently converted by GAME1 to T-Gal (Itkin et al., 2011 supra). T-Gal in its turn is glucosylated by GAME17 into γ-tomatine, which is further glucosylated by GAME18 to β1-tomatine that is finally converted to α-tomatine by GAME2 (**FIGURE 1**).

[00263] As described herein, by modifying expression of an enzyme and/or other protein involved in the biosynthetic pathway, the level of steroidal alkaloids, steroidal glycoalkaloids and steroidal saponin can be altered.

[00264] Silencing of a single gene co-expressed with the clustered enzyme-encoding gene in potato plant, resulted in significant reduction in the amount of the steroidal glycoalkaloids α-chaconine and α-solanine, while overexpression of this gene resulted in significant increase in the content of these substances (**FIGURES 5 and 6**). This gene was found to include coding sequence comprising an AP2 domain, and therefore postulated to be a transcription factor, designated herein GAME9-transcription factor, encoded by GAME9.

[00265] A genetically modified plant comprising at least one cell having altered expression of at least one gene selected from the group consisting of a gene encoding GAME9-transcription factor, a gene encoding 2-oxoglutarate-dependent dioxygenase, a gene encoding basic helix-loop-helix (BHLH)-transcription factor or a combination thereof, wherein the genetically modified plant has an

altered content of at least one steroidal alkaloid or a derivative, metabolite, or biosynthetic intermediate thereof compared to a corresponding unmodified plant, has been produced. As exemplified herein for 2-oxoglutarate-dependent dioxygenase (GAME11), manipulating the expression of the genes of the present invention can further lead to the manipulation of steroidal saponin synthesis.

[00266] Thus, according to additional aspect, provided herein is a genetically modified organism comprising at least one cell having altered expression of at least one gene selected from the group consisting of a gene encoding GAME9-transcription factor, a gene encoding 2-oxoglutarate-dependent dioxygenase, a gene encoding basic helix-loop-helix (BHLH)-transcription factor or a combination thereof compared to an unmodified or unedited organism, wherein the genetically modified organism has an altered content of at least one compound selected from steroidal saponins and/or steroidal alkaloids and their glycosylated and other derivatives thereof, metabolites thereof, or biosynthetic intermediates thereof, compared to a corresponding unmodified or unedited organism.

[00267] Unexpectedly, the present invention now shows that SGA levels can be severely reduced in potato tubers by modifying expression of an enzyme and/or transcription factors involved in the steroidal alkaloid biosynthetic pathway.

[00268] According to certain embodiments, the expression of the at least one gene selected from the group consisting of a gene encoding GAME9-transcription factor, a gene encoding 2-oxoglutarate-dependent dioxygenase, a gene encoding BHLH-transcription factor or the combination thereof in the genetically modified plant is inhibited compared to its expression in the corresponding unmodified or unedited plant, thereby the genetically modified plant comprises reduced content of at least one steroidal alkaloid or a derivative, metabolite, or biosynthetic intermediate thereof, compared to a corresponding unmodified plant.

[00269] The genes and encoded enzymes thereof, active in the spinach triterpenoid biosynthetic pathway of **Compound 11**, are described in detail in the Examples. **Table 17**, provided in **Example 20**, provides the amino acid and nucleic acid sequences of these enzyme and the genes encoding them. **Table 19** provided in **Example 24** provides the amino acid sequence of an additional CSLG enzyme, isolated from *Arabidopsis*.

[00270] In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00271] In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00272] In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00273] In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG enzyme is set forth in a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00274] In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 30. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 32. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 34. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 36. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 38. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 40. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 45. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 46. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 47. In some

embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 51. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 53. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 55. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 57. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 59. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 61. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 163 In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 65. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 80,. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 93. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 95. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 97. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 99. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 101. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 103. In some embodiments, the polynucleotide sequence of said at least one heterologous gene is set forth in SEQ ID NO: 105.

[00275] In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 30. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 32. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 34. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 36. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 38. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 40. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises

a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 45. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 46. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 47. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 51. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 53. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 55. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 57. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 59. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 61. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 63. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 65. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 80. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 93. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 95. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 97. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 99. In some embodiments, the polynucleotide sequence of said at least one

heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 101. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 103. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in SEQ ID NO: 105.

[00276] In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55%, 60%. 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity to the nucleic acid sequence set forth in any of SEQ ID NO: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55%, 60%. 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% coverage to the nucleic acid sequence set forth in any of SEQ ID NO: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00277] In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 30. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 32. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 34. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 36. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 38. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 40. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 45. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 46. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 47. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the

nucleic acid sequence set forth in SEQ ID NO: 51. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 53. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 55. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 57. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 59. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 61. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 63. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 65. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 80. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 93. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 95. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 97. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 99. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 101. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 103. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in SEQ ID NO: 105.

[00278] In some embodiments, the polynucleotide sequence of said at least one heterologous gene

comprises a polynucleotide sequence encoding at least one steroidal alkaloid biosynthetic enzyme, steroidal saponin biosynthetic enzyme, or triterpenoid saponin biosynthetic enzyme. In some embodiments, the polynucleotide sequence of said at least one heterologous gene comprises a polynucleotide sequence encoding 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more at least one steroidal alkaloid biosynthetic enzymes, steroidal saponin biosynthetic enzymes, or triterpenoid saponin biosynthetic enzymes, or combinations thereof.

[00279] A skilled artisan would appreciate that sequence homology encompasses similarity of sequence attributed to descent from a common ancestor. Homologous biological components (genes, proteins, structures) are called homologs. The extent to which nucleotide or protein sequences are related. The similarity between two sequences (DNA, RNA, or amino acid) can be expressed as percent sequence identity and/or percent positive substitutions.

[00280] A skilled artisan would appreciate that the term "homolog" encompasses a gene or a polypeptide (a protein) that is related to a second gene or polypeptide (protein), respectively, by descent from a common ancestral DNA or polypeptide (protein) sequence, respectively. Thus, a homolog of a gene, in some embodiments, comprises a similar nucleotide sequence to the gene. In some embodiments, a gene homolog encodes an identical polypeptide as is encoded by the gene. In some embodiments, a gene homolog encodes a polypeptide with the same functional properties as is encoded by the gene. In some embodiments, a gene homolog encodes a polypeptide that comprises a similar amino acid sequence as the polypeptide encoded by the gene. In one embodiment, the polypeptide homolog comprises a similar amino acid sequence as the polypeptide. In some embodiments, the polypeptide homolog comprises the same functional properties as the polypeptide. In some embodiments, the polypeptide homolog comprises similar functional properties as the polypeptide. In some embodiments, the polypeptide homolog comprises a same domain(s) as the polypeptide. In some embodiments, the polypeptide homolog comprises a similar domain(s) as the polypeptide.

[00281] A skilled artisan would appreciate that percent homology or percent identity may be determined, for example but no limited to, using BlastP software of the National Center of Biotechnology Information (NCBI) using default parameters. The homolog may also refer to an ortholog, a deletion, insertion, or substitution variant, including an amino acid substitution. In some embodiments, sequence identity or homology can be determined using any protein or nucleic acid sequence alignment algorithm such as Blast, ClustalW, MUSCLE, and HHpred.

[00282] A skilled artisan would appreciate that homologs, may in some embodiments, have the same activity. In some embodiments, a SOAP gene homolog has the same activity as the corresponding SOAP gene. In some embodiments, a SOAP polypeptide homolog has the same

enzyme activity as the corresponding SOAP polypeptide. In some embodiments, a SOAP5 homolog has the same enzyme activities as SOAP5. In some embodiments, a GAME gene homolog has the same activity as the corresponding GAME gene. In some embodiments, a GAME polypeptide homolog has the same enzyme activity as the corresponding GAME polypeptide. In some embodiments, a GAME15 homolog has the same enzyme activities as GAME15. In certain embodiments, a homolog of a CSLG gene encodes an enzyme having cellulose synthase like G activity. In certain embodiments, a homolog of a CSLG gene encodes an enzyme having glucuronic acid transferase activity. In certain embodiments, a homolog of a CSLG gene encodes an enzyme having cellulose synthase like G activity and glucuronic acid transferase activity. **Figures 42A** and **42F** provide phylogenetic trees of cellulose synthase like enzymes, wherein the cellulose synthase like G subclade includes the SOAP5 and GAME15 enzymes described herein.

[00283] In some embodiments, a genetically modified cell disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, a genetically modified cell disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00284] In some embodiments, a genetically modified cell disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00285] In some embodiments, a genetically modified cell disclosed herein comprises a nucleic acid sequence encoding an at least one additional heterologous gene in addition to a heterologous CSLG gene, wherein the additional gene encodes

(a) a β-amyrin synthase, and said nucleic acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or

(b) a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or

(c) a glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or

(d) an acetyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or

(e) a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

[00286] In some embodiments, a genetically modified plant disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, a genetically modified plant disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00287] In some embodiments, a genetically modified plant disclosed herein comprises a nucleic acid sequence encoding said at least one heterologous CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00288] In some embodiments, a genetically modified plant disclosed herein comprises a nucleic acid sequence encoding an at least one additional heterologous gene in addition to a heterologous

CSLG gene, wherein the additional gene encodes

(a) a β-amyrin synthase, and said nucleic acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or

(b) a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or

(c) a glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or

(d) an acetyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or

(e) a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

[00289] In some embodiments, a polynucleotide sequence comprising an at least one heterologous gene encoding an enzyme disclosed here, comprises a nucleic acid sequence having at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% homology to a sequence selected from 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a combination thereof. In some embodiments, the nucleic acid sequence of a heterologous gene encoding an enzyme disclosed herein, comprises a nucleic acid sequence having at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identity to a sequence selected from 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a combination thereof.

[00290] In some embodiments, the nucleic acid sequence of a heterologous gene encoding an enzyme disclosed herein, comprises a nucleic acid sequence selected from selected from 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a combination thereof.

[00291] In some embodiments, a nucleic acid sequence having at least 55%, at least 60%, at least

65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identity to a sequence selected from 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, comprises a codon optimized gene sequence.

[00292] In some embodiments, optimizing a gene entails adapting the codon usage to the codon bias of host genes, which in one embodiment comprises a plant host. In some embodiments, optimizing a gene entails adapting the codon usage to the codon bias of host genes, which in one embodiment comprises a yeast host. In some embodiments, optimizing a gene entails adapting the codon usage to the codon bias of host genes, which in one embodiment comprises a algal host. Codon optimized gene sequences may in some embodiments, improve protein expression level in living organism by increasing translational efficiency of an at least one heterologous gene as described herein. In some embodiments, codon-optimization comprising codon optimization for expression in yeast. In some embodiments, codon-optimization comprising codon optimization for expression in algae. In some embodiments, codon-optimization comprising codon optimization for expression in a plant cell.

[00293] In some embodiments, the term "optimized" encompasses a desired change, which, in one embodiment, is a change in gene expression and, in another embodiment, in protein expression. In one embodiment, optimized gene expression is optimized regulation of gene expression. In another embodiment, optimized gene expression is an increase in gene expression. According to this aspect and in one embodiment, a 2-fold through 1000-fold increase in gene expression compared to wild-type is contemplated. In another embodiment, a 2-fold to 500-fold increase in gene expression, in another embodiment, a 2-fold to 100-fold increase in gene expression, in another embodiment, a 2-fold to 50-fold increase in gene expression, in another embodiment, a 2-fold to 20-fold increase in gene expression, in another embodiment, a 2-fold to 10-fold increase in gene expression, in another embodiment, a 3-fold to 5-fold increase in gene expression is contemplated.

[00294] In another embodiment, optimized gene expression may be an increase in gene expression under particular environmental conditions. In another embodiment, optimized gene expression may comprise a decrease in gene expression, which, in one embodiment, may be only under particular environmental conditions.

[00295] In another embodiment, optimized gene expression is an increased duration of gene expression. According to this aspect and in one embodiment, a 2-fold through 1000-fold increase in the duration of gene expression compared to wild-type is contemplated. In another embodiment, a 2-fold to 500-fold increase in the duration of gene expression, in another embodiment, a 2-fold to 100-fold increase in the duration of gene expression, in another embodiment, a 2-fold to 50-fold increase in the duration of gene expression, in another embodiment, a 2-fold to 20-fold increase in the duration

of gene expression, in another embodiment, a 2-fold to 10-fold increase in the duration of gene expression, in another embodiment, a 3-fold to 5-fold increase in the duration of gene expression is contemplated. In another embodiment, the increased duration of gene expression is compared to gene expression in non-vector-expressing controls, or alternatively, compared to gene expression in wild-type-vector-expressing controls.

[00296] A skilled artisan would appreciate that in some embodiments, more than one heterologous gene encoding an enzyme is expressed in a genetically modified cell. In some embodiments, the cell comprises a polynucleotide sequence encoding more than one heterologous gene. In other embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified cell, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein a polynucleotide sequence may comprise one or more genes. In some embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified cell, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein each polynucleotide sequence comprises one a single gene. In some embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified cell, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein each polynucleotide sequence comprises one or more genes.

[00297] As disclosed throughout, in some embodiments a genetically modified cell comprises a plant cell. In some embodiments, a genetically modified plant cell is comprised in a plant or plant part. In other embodiments, a genetically modified cell comprises a yeast cell. In other embodiments a genetically modified cell comprises an algal cell. In other embodiments, a genetically modified cell comprises an insect cell. In other embodiments, a genetically modified cell comprises a bacterium.

[00298] Additionally, a skilled artisan would appreciate that in some embodiments, more than one heterologous gene encoding an enzyme is expressed in a genetically modified plant. In some embodiments, the plant comprises a polynucleotide sequence encoding more than one heterologous gene. In other embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified plant, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein a polynucleotide sequence may comprise one or more genes. In some embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified plant, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein each polynucleotide sequence comprises one a single gene. In some embodiments, when more than one heterologous gene encoding an enzyme is expressed in a genetically modified plant, multiple polynucleotide sequences may be used to encode the multiple heterologous genes, wherein each polynucleotide sequence comprises one or more genes.

[00299] In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene comprises the amino acid sequence of a steroidal alkaloid biosynthetic enzyme, a steroidal saponin biosynthetic enzyme, and/or a triterpenoid biosynthetic enzyme.

[00300] In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 31. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 33. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 35. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 37. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 39. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 41. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 48. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 49. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 50. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 52. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 54. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 56. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 58. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 60. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 62. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 64. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 66. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 69. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 81. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 94. In some embodiments, the amino acid sequence of

an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 96. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 98. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 100. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 102. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in SEQ ID NO: 104.

[00301] In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 31. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 33. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 35. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 37. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 39. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 41. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 48. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 49. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 50. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 52. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity

to the amino acid sequence set forth in SEQ ID NO: 54. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 56. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 58. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 60. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 62. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 64. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 66. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 69. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 81. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 94. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 96. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 98. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 100. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 102. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55% identity to the amino acid sequence set forth in SEQ ID NO: 104.

[00302] In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55%, 60%, 65%, 70%, 75%, 80%, 85%,

90%, 95%, 98%, or 99% identity to the amino acid sequence set forth in any of SEQ ID NO: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% coverage to the amino acid sequence set forth in any of SEQ ID NO: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104.

[00303] In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 31. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 33. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 35. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 37. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 39. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 41. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 48. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 49. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 50. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 52. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 54. In some embodiments, the amino acid

sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 56. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 58. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 60. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 62. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 64. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 66. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 69. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 81. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 94. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 96. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 98. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 100. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 102. In some embodiments, the amino acid sequence of an enzyme encoded by an at least one heterologous gene is set forth in a sequence having at least 80% identity to the amino acid sequence set forth in SEQ ID NO: 104.

[00304] In some embodiments, the enzyme encoded by an at least one heterologous gene comprises an amino acid sequence having at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% homology to a sequence selected from SEQ ID

NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 66, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the enzyme encoded by the at least one heterologous gene comprises the amino acid sequence having at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 95% identity with a sequence selected from SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 66, 81, 94, 96, 98, 100, 102 or 104.

[00305] In some embodiments, the enzyme encoded by an at least one heterologous gene comprises an amino acid sequence having at least 80%, at least 85%, at least 90%, or at least 95% homology to a sequence selected from SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 66, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the enzyme encoded by the at least one heterologous gene comprises the amino acid sequence having at least 80%, at least 85%, at least 95% identity with a sequence selected from SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 66, 81, 94, 96, 98, 100, 102 or 104.

[00306] In some embodiments, the amino acid sequence of the encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, the amino acid sequence of the encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104.

[00307] In some embodiments, the amino acid sequence of the encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104.

[00308] In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes a β-amyrin synthase, said amino acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 45. In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes a cytochrome P450, said amino acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53. In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes a glycosyl transferase, said amino acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80%

coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61. In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes an acetyltransferase, said amino acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 63. In some embodiments, the amino acid sequence of said encoded at least one additional heterologous gene encodes a UDP-glucose 6-dehydrogenase 1, said amino acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 74.

[00309] In some embodiments, a genetically modified plant cell has an altered expression of at least the endogenous CSLG gene compared to the expression of the endogenous CSLG in a corresponding unmodified cell. In some embodiments, a genetically modified plant cell has an altered expression of at least the endogenous CSLG gene compared to the expression of the endogenous CSLG in a corresponding unmodified cell, wherein the cell has an altered content of at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof.

[00310] Altering the expression of the endogenous CSLG affects the steroidal alkaloid biosynthetic pathway, the steroidal saponin biosynthetic pathway, and/or the triterpenoid saponin biosynthetic pathway, and in some embodiments, results in concomitant alteration, respectively, in the steroidal alkaloid profile or intermediates of that pathway, in the steroidal saponin profile or intermediates of that pathway, or in the triterpenoid saponin profile or intermediates of that pathway.

[00311] In some embodiments, the altered expression in a genetically modified cells comprises increased CSLG expression compared with a corresponding unmodified cell. In some embodiments, increased expression results in an increased amount of the encoded enzyme. In some embodiments, increased expression of CSLG results in the alteration in the steroidal alkaloid profile comprising an increase in at least one steroidal alkaloid or an intermediate thereof. In some embodiments, increased expression of CSLG results in the alteration in the steroidal saponin profile comprising an increase in at least one steroidal saponin or an intermediate thereof. In some embodiments, increased expression of CSLG results in the alteration in the triterpenoid saponin profile comprising an increase in at least one triterpenoid saponin or an intermediate thereof.

[00312] In some embodiments, the altered expression in a genetically modified cells comprises decreased CSLG expression compared with a corresponding unmodified cell. In some embodiments, decreased expression results in a decreased amount of the encoded enzyme. In some embodiments, decreased expression of CSLG results in the alteration in the steroidal alkaloid profile comprising a decrease in at least one steroidal alkaloid or an intermediate thereof. In some embodiments, decreased expression of CSLG results in the alteration in the steroidal saponin profile comprising a

decrease in at least one steroidal saponin or an intermediate thereof. In some embodiments, decreased expression of CSLG results in the alteration in the triterpenoid saponin profile comprising a decrease in at least one triterpenoid saponin or an intermediate thereof.

[00313] In some embodiments, the at least one cell having an altered endogenous CSLG gene expression comprises a plant cell. In some embodiments, the at least one cell having an altered endogenous CSLG gene expression comprises a plant cell comprised within a plant or plant part. In some embodiments, the at least one cell having an altered endogenous CSLG gene expression is a result of a mutation in said CSLG gene. In some embodiments, said mutation comprises at least one or more point mutations, or an insertion, or a deletion, or any combination thereof, wherein said expressed CSLG enzyme has increased stability or increased activity or both and the altered content comprises an increased amount of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant. In some embodiments, said mutation comprises at least one or more point mutations, or an insertion, or a deletion, or any combination thereof, wherein said expressed CSLG enzyme has decreased stability or decreased activity or both and the altered content comprises a decreased amount of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[00314] According to certain embodiments, expression of the endogenous CSLG gene is altered, the altering comprising mutagenizing the CSLG gene, wherein the mutagenesis comprises introduction of one or more point mutations, insertions, deletions, or genome editing, or use of a bacterial CRISPR/CAS system, or a combination thereof. According to certain embodiments, expression of the gene encoding the CLSG enzyme is increased compared to its expression in the corresponding unmodified plant, thereby the genetically modified plant comprises increased content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; increased content of at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or increased content of at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to the corresponding unmodified plant. According to certain embodiments, expression of the gene encoding the CLSG enzyme is decreased compared to its expression in the corresponding unmodified plant, thereby the genetically modified plant comprises decreased content at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; decreased content at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or decreased content at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00315] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00316] Inserting a mutation to the CSLG gene, including deletions, insertions, site specific mutations, zinc-finger nucleases and the like can be used for example for upregulation of gene expression, increased stability of the expressed CSLG enzyme, increased activity of the CSLG enzyme, down-regulation of the gene expression, decreased stability of the expressed CSLG enzyme, decreased activity of the CSLG enzyme, or elimination of activity of the CSLG enzyme.

[00317] According to certain embodiments, a mutated CSLG gene comprises a polynucleotide having at least one mutation in the nucleic acid sequence of a CSLG gene, wherein the nucleic acid sequence of said CSLG is set forth in any one SEQ ID NOS: 31, 33, 35, 37, 39, 41, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a complementary sequence thereof.

[00318] One of ordinary skill in the art would appreciate that a genetically modified plant may encompass a plant comprising at least one cell genetically modified by man. In some embodiments, the genetic modification includes modification of an endogenous gene(s), for example by introducing mutation(s) deletions, insertions, transposable element(s) and the like into an endogenous polynucleotide or gene of interest. Additionally, or alternatively, in some embodiments, the genetic modification includes transforming at least one plant cell with a heterologous polynucleotide or multiple heterologous polynucleotides. The skilled artisan would appreciate that a genetically modified plant comprising transforming at least one plant cell with a heterologous polynucleotide or multiple heterologous polynucleotides may in certain embodiments be termed a “transgenic plant” or a “genetically modified plant” having the same meanings and qualities.

[00319] Altering the expression of an endogenous CSLG gene may in certain embodiments be achieved by the introduction of one or more point mutations into a nucleic acid molecule encoding the corresponding protein. Mutations can be introduced using, for example, site-directed mutagenesis (see, e.g., Wu Ed., 1993 Meth. In Enzymol. Vol. 217, San Diego: Academic Press; Higuchi, "Recombinant PCR" in Innis et al. Eds., 1990 PCR Protocols, San Diego: Academic Press, Inc). Such mutagenesis can be used to introduce a specific, desired amino acid insertion, deletion or substitution. Several technologies for targeted mutagenesis are based on the targeted induction of double-strand breaks (DSBs) in the genome followed by error-prone DNA repair. Mostly commonly used for genome editing by these methods are custom designed nucleases, including zinc finger nucleases, endonucleases including meganucleases, and *Xanthomonas*-derived transcription activator-like effector nuclease (TALEN) enzymes.

[00320] In some embodiments, when the expression of the CSLG gene is altered, said altering

comprises mutagenizing CSLG gene, said mutation present within a coding region of said CSLG gene, or a regulatory sequence of said CSLG gene, or a combination thereof.

[00321] Various types of mutagenesis can be used to modify *CSLG* and the encoded enzyme thereof in order to produce conservative or non-conservative variants. Any available mutagenesis procedure can be used. In some embodiments, the mutagenesis procedure comprises site-directed point mutagenesis. In some embodiments, the mutagenesis procedure comprises random point mutagenesis. In some embodiments, the mutagenesis procedure comprises *in vitro* or *in vivo* homologous recombination (DNA shuffling). In some embodiments, the mutagenesis procedure comprises mutagenesis using uracil-containing templates. In some embodiments, the mutagenesis procedure comprises oligonucleotide-directed mutagenesis. In some embodiments, the mutagenesis procedure comprises phosphorothioate-modified DNA mutagenesis. In some embodiments, the mutagenesis procedure comprises mutagenesis using gapped duplex DNA. In some embodiments, the mutagenesis procedure comprises point mismatch repair. In some embodiments, the mutagenesis procedure comprises mutagenesis using repair-deficient host strains. In some embodiments, the mutagenesis procedure comprises restriction-selection and restriction-purification. In some embodiments, the mutagenesis procedure comprises deletion mutagenesis. In some embodiments, the mutagenesis procedure comprises mutagenesis by total gene synthesis. In some embodiments, the mutagenesis procedure comprises double-strand break repair. In some embodiments, the mutagenesis procedure comprises mutagenesis by chimeric constructs. In some embodiments, the mutagenesis procedure comprises mutagenesis by CRISPR/Cas. In some embodiments, the mutagnesis procedure comprises mutagenesis by a meganuclease. In some embodiments, the mutagenesis procedure comprises mutagenesis by zinc-finger nucleases (ZFN). In some embodiments, the mutagenesis procedure comprises mutagenesis by transcription activator-like effector nucleases (TALEN). In some embodiments, the mutagenesis procedure comprises any other mutagenesis procedure known to a person skilled in the art.

[00322] In some embodiments, mutagenesis can be guided by known information about the naturally occurring molecule and/or the mutated molecule. By way of example, this known information may include sequence, sequence comparisons, physical properties, crystal structure and the like. In some embodiments, the mutagenesis is essentially random. In some embodiments the mutagenesis procedure is DNA shuffling.

[00323] A skilled artisan would appreciate that clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein (Cas) system comprises genome engineering tools based on the bacterial CRISPR/Cas prokaryotic adaptive immune system. This RNA-based technology is very specific and allows targeted cleavage of genomic DNA guided by a customizable

small noncoding RNA, resulting in gene modifications by both non-homologous end joining (NHEJ) and homology-directed repair (HDR) mechanisms (Belhaj K. et al., 2013. Plant Methods 2013, 9:39). In some embodiments, a CRISPR/Cas system comprises a CRISPR/Cas9 system.

[00324] In some embodiments, a CRISPR/Cas system comprises a single-guide RNA (sgRNA) and/or a Cas protein known in the art. In some embodiments, a CRISPR/Cas system comprises a single-guide RNA (sgRNA) and/or a Cas protein newly created to cleave at a preselected site. The skilled artisan would appreciate that the terms "single-guide RNA", "sgRNA", and "gRNA" are interchangeable having all the same qualities and meanings, wherein an sgRNA may encompass a chimeric RNA molecule which is composed of a CRISPR RNA (crRNA) and trans-encoded CRISPR RNA (tracrRNA). In some embodiments, a crRNA is complementary to a preselected region of *GAME15* DNA, wherein the crRNA "targets" the CRISPR associated polypeptide (Cas) nuclease protein to the preselected target site.

[00325] In some embodiments, the length of crRNA sequence complementary is 19-22 nucleotides long e.g., 19-22 consecutive nucleotides complementary to the target site. In another embodiment, the length of crRNA sequence complementary to the region of DNA is about 15-30 nucleotides long. In another embodiment, the length of crRNA sequence complementary to the region of DNA is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides long. In another embodiment, the length of crRNA sequence complementary to the region of DNA is 20 nucleotides long. In some embodiments, the crRNA is located at the 5' end of the sgRNA molecule. In another embodiment, the crRNA comprises 100% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 80% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 85% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 90% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 95% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 97% complementation within the preselected target sequence. In another embodiment, the crRNA comprises at least 99% complementation within the preselected target sequence. In another embodiment, a tracrRNA is 100-300 nucleotides long and provides a binding site for the Cas nuclease, e.g., a Cas9 protein forming the CRISPR/Cas9 complex.

[00326] In one embodiment, a mutagenesis system comprises a CRISPR/Cas system. In another embodiment, a CRISPR/Cas system comprises a Cas nuclease and a gRNA molecule, wherein said gRNA molecule binds within said preselected endogenous target site thereby guiding said Cas nuclease to cleave the DNA within said preselected endogenous target site.

[00327] In some embodiments, a CRISPR/Cas system comprise an enzyme system including a

guide RNA sequence ("gRNA" or "sgRNA") that contains a nucleotide sequence complementary or substantially complementary to a region of a target polynucleotide, for example a preselected endogenous target site, and a protein with nuclease activity.

[00328] In another embodiment, a CRISPR/Cas system comprises a Type I CRISPR-Cas system, or a Type II CRISPR-Cas system, or a Type III CRISPR-Cas system, or derivatives thereof. In another embodiment, a CRISPR-Cas system comprises an engineered and/or programmed nuclease system derived from naturally accruing CRISPR-Cas systems. In another embodiment, a CRISPR-Cas system comprises engineered and/or mutated Cas proteins. In another embodiment, a CRISPR-Cas system comprises engineered and/or programmed guide RNA.

[00329] A skilled artisan would appreciate that a guide RNA may contain nucleotide sequences other than the region complementary or substantially complementary to a region of a target DNA sequence, for example a preselected endogenous target site. In another embodiment, a guide RNA comprises a crRNA or a derivative thereof. In another embodiment, a guide RNA comprises a crRNA: tracrRNA chimera.

[00330] In another embodiment, a gRNA molecule comprises a domain that is complementary to and binds to a preselected endogenous target site on at least one homologous chromosome. In another embodiment, a gRNA molecule comprises a domain that is complementary to and binds to a polymorphic allele on at least one homologous chromosome. In another embodiment, a gRNA molecule comprises a domain that is complementary to and binds to a preselected endogenous target site on both homologous chromosomes. In another embodiment, a gRNA molecule comprises a domain that is complementary to and binds to polymorphic alleles on both homologous chromosomes.

[00331] Cas enzymes comprise RNA-guided DNA endonuclease able to make double-stranded breaks (DSB) in DNA. The term "Cas enzyme" may be used interchangeably with the terms "CRISPR-associated endonucleases" or "CRISPR-associated polypeptides" having all the same qualities and meanings. In one embodiment, a Cas enzyme is selected from the group comprising Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, C2c1, CasX, NgAgo, Cpf1, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4, or homologs thereof, or modified versions thereof. In another embodiment, a Cas enzyme comprises Cas9. In another embodiment, a Cas enzyme comprises Cas1. In another embodiment, a Cas enzyme comprises Cas1B. In another embodiment, a Cas enzyme comprises Cas2. In another embodiment, a Cas enzyme comprises Cas3. In another embodiment, a Cas enzyme comprises Cas4. In another embodiment, a Cas enzyme comprises Cas5. In another

embodiment, a Cas enzyme comprises Cas6. In another embodiment, a Cas enzyme comprises Cas7. In another embodiment, a Cas enzyme comprises Cas8. In another embodiment, a Cas enzyme comprises Cas10. In another embodiment, a Cas enzyme comprises Cpf1. In another embodiment, a Cas enzyme comprises Csy1. In another embodiment, a Cas enzyme comprises Csy2. In another embodiment, a Cas enzyme comprises Csy3. In another embodiment, a Cas enzyme comprises Cse1. In another embodiment, a Cas enzyme comprises Cse2. In another embodiment, a Cas enzyme comprises Csc1. In another embodiment, a Cas enzyme comprises Csc2. In another embodiment, a Cas enzyme comprises Csa5. In another embodiment, a Cas enzyme comprises Csn2. In another embodiment, a Cas enzyme comprises Csm2. In another embodiment, a Cas enzyme comprises Csm3. In another embodiment, a Cas enzyme comprises Csm4. In another embodiment, a Cas enzyme comprises Csm5. In another embodiment, a Cas enzyme comprises Csm6. In another embodiment, a Cas enzyme comprises Cmr1. In another embodiment, a Cas enzyme comprises Cmr3. In another embodiment, a Cas enzyme comprises Cmr4. In another embodiment, a Cas enzyme comprises Cmr5. In another embodiment, a Cas enzyme comprises Cmr6. In another embodiment, a Cas enzyme comprises Csb1. In another embodiment, a Cas enzyme comprises Csb2. In another embodiment, a Cas enzyme comprises Csb3. In another embodiment, a Cas enzyme comprises Csx17. In another embodiment, a Cas enzyme comprises Csx14. In another embodiment, a Cas enzyme comprises Csx10. In another embodiment, a Cas enzyme comprises Csx16, CsaX. In another embodiment, a Cas enzyme comprises Csx3. In another embodiment, a Cas enzyme comprises Csx1, Csx15, Csf1. In another embodiment, a Cas enzyme comprises Csf2. In another embodiment, a Cas enzyme comprises Csf3. In another embodiment, a Cas enzyme comprises Csf4. In another embodiment, a Cas enzyme comprises Cpf1. In another embodiment, a Cas enzyme comprises C2cl. In another embodiment, a Cas enzyme comprises CasX. In another embodiment, a Cas enzyme comprises NgAgo. In another embodiment, a Cas enzyme is Cas homologue. In another embodiment, a Cas enzyme is a Cas orthologue. In another embodiment, a Cas enzyme is a modified Cas enzyme. In another embodiment, a Cas enzyme is any CRISPR-associated endonucleases known in the art.

[00332] A skilled artisan would appreciate that the terms “zinc finger nuclease” or “ZFN” are interchangeable having all the same meanings and qualities, wherein a ZFN encompasses a chimeric protein molecule comprising at least one zinc finger DNA binding domain operatively linked to at least one nuclease capable of double-strand cleaving of DNA. In some embodiments, a ZFN system comprises a ZFN known in the art. In some embodiments, a ZFN system comprises a ZFN newly created to cleave a preselected site.

[00333] In some embodiments, a ZFN creates a double-stranded break at a preselected endogenous

target site. In some embodiments, a ZFN comprises a DNA-binding domain and a DNA-cleavage domain, wherein the DNA binding domain is comprised of at least one zinc finger and is operatively linked to a DNA-cleavage domain. In another embodiment, a zinc finger DNA- binding domain is at the N-terminus of the chimeric protein molecule and the DNA- cleavage domain is located at the C-terminus of the molecule. In another embodiment, a zinc finger DNA-binding domain is at the C-terminus of the chimeric protein molecule and the DNA- cleavage domain is located at the N-terminus of the molecule. In another embodiment, a zinc finger binding domain encompasses the region in a zinc finger nuclease that is capable of binding to a target locus, for example a preselected endogenous target site as disclosed herein. In another embodiment, a zinc finger DNA-binding domain comprises a protein domain that binds to a preselected endogenous target site on at least one homologous chromosome. In another embodiment, a zinc finger DNA-binding domain comprises a protein domain that binds to a polymorphic allele on at least one homologous chromosome. In another embodiment, a zinc finger DNA-binding domain comprises a protein domain that binds to a preselected endogenous target site on both homologous chromosomes. In another embodiment, a zinc finger DNA-binding domain comprises a protein domain that binds to polymorphic alleles on both homologous chromosomes.

[00334] The skilled artisan would appreciate that the term "chimeric protein" is used to describe a protein that has been expressed from a DNA molecule that has been created by operatively joining two or more DNA fragments. The DNA fragments may be from the same species, or they may be from a different species. The DNA fragments may be from the same or a different gene. The skilled artisan would appreciate that the term "DNA cleavage domain" of a ZFN encompasses the region in the zinc finger nuclease that is capable of breaking down the chemical bonds between nucleic acids in a nucleotide chain. Examples of proteins containing cleavage domains include restriction enzymes, topoisomerases, recombinases, integrases and DNAses.

[00335] A skilled artisan would appreciate that endonucleases include meganucleases, also known as homing endonucleases (HEases), which like restriction endonucleases, bind and cut at a specific DSB target site, however the DSB target sites for meganucleases are typically longer, about 18 bp or more. Meganuclease domains, structure and function are known, see for example, Guhan and Muniyappa (2003) Crit Rev Biochem Mol Biol 38:199-248; Lucas et al., (2001) Nucleic Acids Res 29:960-9; Jurica and Stoddard, (1999) Cell Mol Life Sci 55:1304-26; Stoddard, (2006) Q Rev Biophys 38:49-95; and Moure et al., (2002) Nat Struct Biol 9:764. In some examples a naturally occurring variant, and/or engineered derivative meganudease is used. Any meganudease can be used herein, including, but not limited to, l-Scel, l-Scell, l-Scelll, l-ScelV, l-SceV, l-SceVI, l-SceVII, l-Ceul, l-CeuAIIP, l-Crel, I- CrepsblP, l-CrepsbllP, l-CrepsblllP, l-CrepsblVP, l-Tlil, l-Ppol, Pl-Pspl,

F-Scel, F- Scell, F-Suvl, F-Tevl, F-Tevll, l-Amal, l-Anil, l-Chul, l-Cmoel, l-Cpal, l-Cpall, I- Csml, l-Cvul, l-CvuAIP, l-Ddil, l-Ddill, l-Dirl, l-Dmol, l-Hmul, l-Hmull, l-HsNIP, l-Llal, l-Msol, l-Naal, l-Nanl, l-NcllP, l-NgrIP, l-Nitl, l-Njal, l-Nsp236IP, l-Pakl, l-PbolP, I- PculP, l-PcuAI, l-PcuVI, l-PgrIP, l-PoblP, l-Porl, l-PorllP, l-PbpIP, l-SpBetalP, I- Scal, l-SexlP, l-SnelP, l-Spoml, l-SpomCP, l-SpomlP, l-SpomllP, l-SqulP, I- Ssp6803l, l-SthPhiJP, l-SthPhiST3P, l-SthPhiSTe3bP, l-TdelP, l-Tevl, l-Tevll, I- Tevlll, l-UarAP, l-UarHGPAIP, l-UarHGPA13P, l-VinIP, l-ZbilP, Pl-Mtul, PI-MtuHIP PI-MtuHIIP, Pl-Pful, Pl-Pfull, Pl-Pkol, Pl-Pkoll, PI-Rma43812IP, PI-SpBetalP, Pl- Scel, Pl-Tful, Pl-Tfull, Pl-Thyl, Pl-Tlil, Pl-Tlill, or any active variants or fragments thereof. TAL effector nucleases can be used to make double-strand breaks at specific target sequences in the genome of a plant or other organism. TAL effector nucleases can be created by fusing a native or engineered transcription activatorlike (TAL) effector, or functional part thereof, to the catalytic domain of an endonuclease, such as, for example, FokI. The unique, modular TAL effector DNA binding domain allows for the design of proteins with potentially any given DNA recognition specificity. Thus, the DNA binding domains of the TAL effector nucleases can be engineered to recognize specific DNA target sites and thus, used to make double-strand breaks at desired target sequences. See, WO 2010/079430; Morbitzer et al. (2010) PNAS 10.1073/pnas.1013133107; Scholze & Boch (2010) Virulence 1 :428-432; Christian et al. Genetics (2010) 186:757-761; Li et al. (2010) Nuc. Acids Res. (2010) doi:10.1093/nar/gkq704; and Miller et al. (2011) Nature Biotechnology 29:143-148; all of which are herein incorporated by reference.

[00336] In some embodiments, a TALEN system comprises a TAL effector DNA binding domain and a DNA cleavage domain, wherein said TAL effector DNA binding domain binds within said preselected endogenous target site, thereby targeting the DNA cleavage domain to cleave the DNA within said preselected endogenous target site.

[00337] A skilled artisan would appreciate that the terms "transcription activator-like effector nuclease", "TALEN", and "TAL effector nuclease" may be used interchangeably having all the same meanings and qualities, wherein a TALEN encompasses a nuclease capable of recognizing and cleaving its target site, for example a preselected endogenous target site as disclosed herein. In another embodiment, a TALEN comprises a fusion protein comprising a TALE domain and a nucleotide cleavage domain. In another embodiment, a TALE domain comprises a protein domain that binds to a nucleotide in a sequence-specific manner through one or more TALE-repeat modules. A skilled artisan would recognize that TALE-repeat modules comprise a variable number of about 34 amino acid repeats that recognize plant DNA sequences. Further, repeat modules can be rearranged according to a simple cipher to target new DNA sequences. In another embodiment, a TALE domain comprises a protein domain that binds to a preselected endogenous target site on at

least one homologous chromosome. In another embodiment, a TALE domain comprises a protein domain that binds to a polymorphic allele on at least one homologous chromosome. In another embodiment, a TALE domain comprises a protein domain that binds to a preselected endogenous target site on both homologous chromosomes. In another embodiment, a TALE domain comprises a protein domain that binds to polymorphic alleles on both homologous chromosomes.

[00338] In one embodiment, a TALE domain comprises at least one of the TALE-repeat modules. In another embodiment, a TALE domain comprises from one to thirty TALE-repeat modules. In another embodiment, a TALE domain comprises more than thirty repeat modules. In another embodiment, a TALEN fusion protein comprises an N-terminal domain, one or more of TALE-repeat modules followed by a half-repeat module, a linker, and a nucleotide cleavage domain.

[00339] Chemical mutagenesis using an agent such as Ethyl Methyl Sulfonate (EMS) can be employed to obtain a population of point mutations and screen for mutants of the *CSGL* gene that may become silent or down-regulated. In plants, methods relaying on introgression of genes from natural populations can be used. Cultured and wild types species are crossed repetitively such that a plant comprising a given segment of the wild genome is isolated. Certain plant species, for example, maize (corn) and snapdragon, have natural transposons. These transposons are either autonomous, i.e. the transposase is located within the transposon sequence or non-autonomous, without a transposase. A skilled person can cause transposons to "jump" and create mutations. Alternatively, a nucleic acid sequence can be synthesized having random nucleotides at one or more predetermined positions to generate random amino acid substituting.

[00340] In some embodiments, the expression of an endogenous *CSLG* gene can be altered by the introduction of one or more point mutations into their regulatory sequences. In some embodiments, the expression of a heterologous CSLG gene can be altered by the introduction of one or more point mutations into their regulatory sequences.

[00341] A skilled artisan would appreciate that "regulatory sequences" refers to nucleotide sequences located upstream (5′ non-coding sequences), within, or downstream (3′ non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. In some embodiments, regulatory sequences comprise promoters. In some embodiments, regulatory sequences comprise translation leader sequences. In some embodiments, regulatory sequences comprise introns. In some embodiments, regulatory sequences comprise polyadenylation recognition sequences. In some embodiments, regulatory sequences comprise RNA processing sites. In some embodiments, regulatory sequences comprise effector binding sites. In some embodiments, regulatory sequences comprise stem-loop structures.

[00342] A skilled artisan would appreciate that "promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In some embodiments, a coding sequence is located 3′ to a promoter sequence. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. In some embodiments, the promoter comprises a constitutive promoter, i.e., a promoter that causes a gene to be expressed in most cell types at most times. In some embodiments, the promoter comprises a regulated promoter, i.e., a promoter that causes a gene to be expressed in response to sporadic specific stimuli. It is further recognized that in many cases the exact boundaries of regulatory sequences have not been completely defined yet.

[00343] A skilled artisan would appreciate that the term "3′ non-coding sequences" or "transcription terminator" refers to DNA sequences located downstream of a coding sequence. In some embodiments, 3′ non-coding sequences comprise polyadenylation recognition sequences. In some embodiments, 3′ non-coding sequences comprise sequences encoding regulatory signals capable of affecting mRNA processing. In some embodiments, 3′ non-coding sequences comprise sequences encoding regulatory signals capable of affecting gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3′ end of the mRNA precursor. In some embodiments, mutations in the 3′ non-coding sequences affect gene transcription. In some embodiments, mutations in the 3′ non-coding sequences affect RNA processing. In some embodiments, mutations in the 3′ non-coding sequences affect gene stability. In some embodiments, mutations in the 3′ non-coding sequences affect translation of the associated coding sequence.

[00344] According to certain embodiments, expression of the gene encoding the CSLG enzyme is reduced compared to its expression in the corresponding unmodified plant, thereby the genetically modified plant comprises reduced content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate; reduced content of at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate; or reduced content of at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate, compared to the corresponding unmodified plant.

[00345] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00346] According to certain embodiments, the genetically modified plant is a transgenic plant comprising at least one cell comprising at least one silencing molecule targeted to a polynucleotide

encoding a CLSG enzyme. According to certain embodiments, the transgenic plant comprises a polynucleotide encoding a CLSG enzyme, wherein expression of the polynucleotide expression is selectively silenced, repressed, or reduced. According to certain embodiments, the transgenic plant comprises a polynucleotide encoding a CSLG enzyme, wherein the polynucleotide has been selectively edited by deletion, insertion, or modification to silence, repress, or reduce expression thereof, or wherein the genetically modified plant is a progeny of the gene edited plant.

[00347] According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to a CSLG gene.

[00348] In some embodiments, said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression and said altered content comprises a reduced amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a reduced amount of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a reduced amount of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant. wherein when said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression, said cell further comprises at least one silencing molecule targeted to the polynucleotide encoding said CSLG gene, wherein the silencing molecule is selected from an RNA interference molecule or an antisense molecule, or wherein the silencing molecule is a component of a viral induced gene silencing system.

[00349] According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00350] According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 30. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 32. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 34. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 36. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 38. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID

NO: 40. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 65. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 80. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 93. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 95. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 97. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 99. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 101. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 103. According to certain embodiments, the transgenic plant comprises at least one cell comprising at least one silencing molecule targeted to the nucleic acid sequence set forth in SEQ ID NO: 105.

[00351] According to certain embodiments, the silencing molecule is selected from the group consisting of an RNA interference molecule and an antisense molecule, or wherein the silencing molecule is a component of a viral induced gene silencing system. According to certain embodiments, the silencing molecule comprises a polynucleotide having a nucleic acid sequence substantially complementary to a region of a CSLG gene having the nucleic acid sequence set forth in any one SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a complementary sequence thereof. In some embodiments, an antisense molecule silencing, repressing, or reducing the expression of CSLG in a plant or plant cell comprises a polynucleotide having a nucleic acid sequence substantially complementary to a region of a CSLG gene having the nucleic acid sequence set forth in any one SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a complementary sequence thereof. In some embodiments, an antisense molecule silencing, repressing, or reducing the expression of CSLG in a plant or plant cell has the nucleotide sequence set forth in SEQ ID NO: 108. In some embodiments, a VIGs molecule silencing, repressing, or reducing the expression of CSLG in a plant or plant cell has the nucleotide sequence set forth in SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 106 or SEQ ID NO: 107. According

to certain embodiments, the silencing molecule is targeted to a CSLG fragment having the nucleic acid sequence within the nucleic acid sequence set forth in any one SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105, or a complementary sequence thereof. In some embodiments, the genetically modified plant comprising said plant cell has a reduced content of a triterpenoid saponin, a metabolite thereof, a derivative thereof, or a biosynthetic intermediate thereof.

[00352] According to certain embodiments, the genetically modified or gene edited plant comprise at least one cell comprising at least one silencing molecule targeted to a *CSLG* gene. According to some embodiments, the at least one silencing molecule is selected from the group consisting of RNA interference molecule and antisense molecule. According to these embodiments, the genetically modified plant comprises reduced content of a steroidal alkaloid or a biosynthetic intermediate thereof, a steroidal saponin or a biosynthetic intermediate thereof, or a triterpenoid saponin or a biosynthetic intermediate thereof.

[00353] The silencing molecule target to a CSLG gene can be designed as is known to a person skilled in the art. According to certain embodiments, the silencing molecule comprises a polynucleotide having a nucleic acid sequence substantially complementary to a region of a CSLG gene, for example, but not limited to *GAME15* or homologs thereof disclosed herein in tomato, wild tomato, potato, wild potato, and eggplant, or to a complementary sequence of the *GAME15* gene. According to certain embodiments, the silencing molecule comprises a polynucleotide having a nucleic acid sequence substantially complementary to a region of a CSLG gene, for example, but not limited to *SOAP5* or homologs thereof disclosed herein in Chinese licorice, *Arabidopsis,* red beet, quinoa, alfalfa, soybean, or *Lotus japonicus*, or to a complementary sequence of the *SOAP5* gene.

[00354] According to certain embodiments, the silencing molecule is an antisense RNA.

[00355] According to certain exemplary embodiments, the silencing molecule is an RNA interference (RNAi) molecule. According to some embodiments, the silencing molecule is a double-stranded (ds)RNA molecule. According to certain embodiments, the first and the second polynucleotides are separated by a spacer. According to some embodiments, the spacer sequence is an intron. According to yet further embodiments, the expression of the first and the second polynucleotides is derived from one promoter. According to other embodiments, expression of the first and the second polynucleotides are derived from two promoters; the promoters can be identical or different.

[00356] In some embodiments, the term "RNA interference" or "RNAi" encompasses the silencing or decreasing of gene expression mediated by small double stranded RNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by inhibitory RNA (iRNA) that is homologous in its duplex region to the sequence of the silenced gene. The gene

may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

[00357] Typically, the term RNAi molecule refers to single- or double-stranded RNA molecules comprising both a sense and antisense sequence. For example, the RNA interference molecule can be a double-stranded polynucleotide molecule comprising self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule. Alternatively the RNAi molecule can be a single-stranded hairpin polynucleotide having self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule or it can be a circular single-stranded polynucleotide having two or more loop structures and a stem comprising self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule, and wherein the circular polynucleotide can be processed either in vivo or in vitro to generate an active molecule capable of mediating RNAi.

[00358] Antisense technology is the process in which an antisense RNA or DNA molecule interacts with a target sense DNA or RNA strand. A sense strand is a 5' to 3' mRNA molecule or DNA molecule. The complementary strand, or mirror strand, to the sense is called an antisense. When an antisense strand interacts with a sense mRNA strand, the double helix is recognized as foreign to the cell and will be degraded, resulting in reduced or absent protein production. Although DNA is already a double stranded molecule, antisense technology can be applied to it, building a triplex formation.

[00359] One skilled in the art would appreciate that the terms "complementary" or "complement thereof" are used herein to encompass the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

[00360] Antisense modulation of cell and/or tissue levels of *CSLG* may be effected by transforming the cell or tissue with at least one antisense compound, including antisense DNA, antisense RNA, a ribozyme, DNAzyme, a locked nucleic acid (LNA), and an aptamer. In some embodiments, the molecules are chemically modified. In other embodiments, the antisense molecule is antisense DNA or an antisense DNA analog.

[00361] RNA antisense strands can be either catalytic or non-catalytic. The catalytic antisense

strands, also called ribozymes, cleave the RNA molecule at specific sequences. A non-catalytic RNA antisense strand blocks further RNA processing.

[00362] Antisense modulation of cells and/or tissue levels of the CSLG gene may be effected by transforming the organism cells or tissues with at least one antisense compound, including antisense DNA, antisense RNA, a ribozyme, DNAzyme, a locked nucleic acid (LNA), a VIGs, and an aptamer. In some embodiments the molecules are chemically modified. In other embodiments, the antisense molecule is antisense DNA or an antisense DNA analog.

[00363] RNAi refers to the introduction of homologous double stranded RNA (dsRNA) to target a specific gene product, resulting in post transcriptional silencing of that gene. This phenomenon was first reported in *Caenorhabditis elegans* by Guo and Kemphues (1995, Cell, 81(4):611-620) and subsequently Fire et al. (1998, Nature 391:806-811) discovered that it is the presence of dsRNA, formed from the annealing of sense and antisense strands present in the in vitro RNA preps, that is responsible for producing the interfering activity

[00364] In both plants and animals, RNAi is mediated by RNA-induced silencing complex (RISC), a sequence-specific, multicomponent nuclease that destroys messenger RNAs homologous to the silencing trigger. RISC is known to contain short RNAs (approximately 22 nucleotides) derived from the double-stranded RNA trigger. The short-nucleotide RNA sequences are homologous to the target gene that is being suppressed. Thus, the short-nucleotide sequences appear to serve as guide sequences to instruct a multicomponent nuclease, RISC, to destroy the specific mRNAs .

[00365] The dsRNA used to initiate RNAi, may be isolated from native source or produced by known means, e.g., transcribed from DNA. Plasmids and vectors for generating RNAi molecules against target sequence are now readily available from commercial sources.

[00366] The dsRNA can be transcribed from the vectors as two separate strands. In other embodiments, the two strands of DNA used to form the dsRNA may belong to the same or two different duplexes in which they each form with a DNA strand of at least partially complementary sequence. When the dsRNA is thus-produced, the DNA sequence to be transcribed is flanked by two promoters, one controlling the transcription of one of the strands, and the other that of the complementary strand. These two promoters may be identical or different. Alternatively, a single promoter can derive the transcription of single-stranded hairpin polynucleotide having self-complementary sense and antisense regions that anneal to produce the dsRNA.

[00367] Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. RNA molecules containing a nucleotide sequence identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found

to be effective for inhibition. Thus, sequence identity may be optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript. The length of the identical nucleotide sequences may be at least 25, 50, 100, 200, 300 or 400 bases. There is no upper limit on the length of the dsRNA that can be used. For example, the dsRNA can range from about 21 base pairs (bp) of the gene to the full length of the gene or more.

[00368] The term "RNA interference" or "RNAi" refers to the silencing or decreasing of gene expression mediated by small double stranded RNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by inhibitory RNA (iRNA) that is homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

[00369] One of ordinary skill in the art would appreciate that the term RNAi molecule refers to single- or double-stranded RNA molecules comprising both a sense and antisense sequence. For example, the RNA interference molecule can be a double-stranded polynucleotide molecule comprising self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule. Alternatively the RNAi molecule can be a single-stranded hairpin polynucleotide having self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule or it can be a circular single-stranded polynucleotide having two or more loop structures and a stem comprising self-complementary sense and antisense regions, wherein the antisense region comprises complementarity to a target nucleic acid molecule, and wherein the circular polynucleotide can be processed either *in vivo* or *in vitro* to generate an active molecule capable of mediating RNAi.

[00370] In some embodiments, the use of RNA interference (RNAi) to down regulate the expression of a *CSLG* gene is to attenuate the level in plants or a cell thereof or part thereof. In both plants and animals, RNAi is mediated by RNA-induced silencing complex (RISC), a sequence-

specific, multicomponent nuclease that destroys messenger RNAs homologous to the silencing trigger. RISC is known to contain short RNAs (approximately 22 nucleotides) derived from the double-stranded RNA trigger. The short-nucleotide RNA sequences are homologous to the target gene that is being suppressed. Thus, the short-nucleotide sequences appear to serve as guide sequences to instruct a multicomponent nuclease, RISC, to destroy the specific mRNAs .

[00371] The dsRNA used to initiate RNAi, may be isolated from native source or produced by known means, e.g., transcribed from DNA. Plasmids and vectors for generating RNAi molecules against target sequence are now readily available as exemplified herein below.

[00372] According to certain exemplary embodiments, the dsRNA is targeted to *CSLG*.

[00373] In some embodiments, the use of RNA interference (RNAi) to down regulate the expression of a *CSLG* gene is to attenuate the level of in plants or a cell thereof or part thereof. In both plants and animals, RNAi is mediated by RNA-induced silencing complex (RISC), a sequence-specific, multicomponent nuclease that destroys messenger RNAs homologous to the silencing trigger. RISC is known to contain short RNAs (approximately 22 nucleotides) derived from the double-stranded RNA trigger. The short-nucleotide RNA sequences are homologous to the target gene that is being suppressed. Thus, the short-nucleotide sequences appear to serve as guide sequences to instruct a multicomponent nuclease, RISC, to destroy the specific mRNAs .

[00374] The dsRNA used to initiate RNAi, may be isolated from native source or produced by known means, e.g., transcribed from DNA. Plasmids and vectors for generating RNAi molecules against target sequence are now readily available as exemplified herein below.

[00375] The dsRNA can be transcribed from the vectors as two separate strands. In other embodiments, the two strands of DNA used to form the dsRNA may belong to the same or two different duplexes in which they each form with a DNA strand of at least partially complementary sequence. When the dsRNA is thus-produced, the DNA sequence to be transcribed is flanked by two promoters, one controlling the transcription of one of the strands, and the other that of the complementary strand. These two promoters may be identical or different. Alternatively, a single promoter can derive the transcription of single-stranded hairpin polynucleotide having self-complementary sense and antisense regions that anneal to produce the dsRNA.

[00376] Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. RNA molecules containing a nucleotide sequence identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer,

Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript. The length of the identical nucleotide sequences may be at least 25, 50, 100, 200, 300 or 400 bases. There is no upper limit on the length of the dsRNA that can be used. For example, the dsRNA can range from about 21 base pairs (bp) of the gene to the full length of the gene or more.

[00377] Another agent capable of down-regulating the expression of *CSLG* is a Co-Suppression molecule. Co-suppression is a post-transcriptional mechanism where both the transgene and the endogenous gene are silenced.

[00378] Another agent capable of down-regulating the expression of *CSLG* is a DNAzyme molecule, which is capable of specifically cleaving an mRNA transcript or a DNA sequence of the *CSLG*. DNAzymes are single-stranded polynucleotides that are capable of cleaving both single- and double-stranded target sequences. A general model (the "10-23" model) for the DNAzyme has been proposed. "10-23" DNAzymes have a catalytic domain of 15 deoxyribonucleotides, flanked by two substrate-recognition domains of seven to nine deoxyribonucleotides each. This type of DNAzyme can effectively cleave its substrate RNA at purine:pyrimidine junctions (for review of DNAzymes, see: Khachigian, L. M. (2002) Curr Opin Mol Ther 4, 119-121).

[00379] Examples of construction and amplification of synthetic, engineered DNAzymes recognizing single- and double-stranded target cleavage sites are disclosed in U.S. Patent No. 6,326,174.

[00380] The terms "enzymatic nucleic acid molecule" or "enzymatic oligonucleotide" refers to a nucleic acid molecule which has complementarity in a substrate binding region to a specified gene target, and also has an enzymatic activity which is active to specifically cleave target RNA of *CSLG*, thereby silencing the gene. The complementary regions allow sufficient hybridization of the enzymatic nucleic acid molecule to the target RNA and subsequent cleavage. The term enzymatic nucleic acid is used interchangeably with for example, ribozymes, catalytic RNA, enzymatic RNA, catalytic DNA, aptazyme or aptamer-binding ribozyme, catalytic oligonucleotide, nucleozyme, DNAzyme, RNAenzyme. The specific enzymatic nucleic acid molecules described in the instant application are not limiting and an enzymatic nucleic acid molecule of disclosed herein requires a specific substrate binding site which is complementary to one or more of the target nucleic acid

regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart a nucleic acid cleaving and/or ligation activity to the molecule. US Patent No. 4,987,071 discloses examples of such molecules.

[00381] According to certain embodiments, the *CSLG* gene or a silencing molecule targeted thereto form part of an expression vector comprising all necessary elements for expression of the gene or its silencing molecule. According to certain embodiments, the expression is controlled by a constitutive promoter. According to certain embodiments, the constitutive promoter is specific to a plant tissue. According to these embodiments, the tissue specific promoter is selected from the group consisting of root, tuber, leaves and fruit specific promoter. Root specific promoters are described, e.g. in Martinez, E. et al. 2003. Curr. Biol. 13:1435-1441. Fruit specific promoters are described among others in Estornell L.H et al. 2009. Plant Biotechnol. J. 7:298-309 and Fernandez A. I. Et al. 2009 Plant Physiol. 151:1729-1740. Tuber specific promoters are described, e.g. in Rocha-Sosa M, et al., 1989. EMBO J. 8:23-29; McKibbin R.S. et al., 2006. Plant Biotechnol J. 4(4):409-18. Leaf specific promoters are described, e.g. in Yutao Yang, Guodong Yang, Shijuan Liu, Xingqi Guo and Chengchao Zheng. Science in China Series C: Life Sciences. 46: 651-660.

[00382] According to certain embodiments, the expression vector further comprises regulatory elements at the 3' non-coding sequence. As used herein, the "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht I L et al. (1989. Plant Cell 1:671-680).

[00383] Those skilled in the art will appreciate that the various components of the nucleic acid sequences and the transformation vectors described herein, are operatively linked, so as to result in expression of said nucleic acid or nucleic acid fragment. Techniques for operatively linking the components of the constructs and vectors disclosed herein are well known to those skilled in the art. Such techniques include the use of linkers, such as synthetic linkers, for example including one or more restriction enzyme sites.

[00384] Detection of a mutated *CSLG* gene and/or the presence of silencing molecule targeted to the gene and/or over-expression of the genes is performed employing standard methods of molecular genetics, known to a person of ordinary skill in the art.

[00385] For measuring the gene(s) or silencing molecule(s) expression, cDNA or mRNA should be obtained from an organ in which the nucleic acid is expressed. The sample may be further processed before the detecting step. For example, the polynucleotides in the cell or tissue sample may

be separated from other components of the sample, may be amplified, etc. All samples obtained from an organism, including those subjected to any sort of further processing are considered to be obtained from the organism.

[00386] Detection of the gene(s) or the silencing molecule(s) typically requires amplification of the polynucleotides taken from the candidate altered organism. Methods for DNA amplification are known to a person skilled in the art. Most commonly used method for DNA amplification is PCR (polymerase chain reaction; see, for example, PCR Basics: from background to Bench, Springer Verlag, 2000; Eckert et al., 1991. PCR Methods and Applications 1:17). Additional suitable amplification methods include the ligase chain reaction (LCR), transcription amplification and self-sustained sequence replication, and nucleic acid-based sequence amplification (NASBA).

[00387] According to certain embodiments, the nucleic acid sequence comprising the *CSLG* gene or its silencing molecule further comprises a nucleic acid sequence encoding a selectable marker. According to certain embodiments, the selectable marker confers resistance to antibiotic or to an herbicide; in these embodiments the transgenic plants are selected according to their resistance to the antibiotic or herbicide.

[00388] In some embodiments, said endogenous CSLG gene is selectively overexpressed and said altered content comprises an increased amount of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

[00389] Overexpression of the CSLG gene can be obtained by any method as is known to a person skilled in the art. According to certain embodiments, disclosed herein are genetically modified plants or parts thereof comprising at least one cell comprising at least one transcribable polynucleotide that is over expressed and encodes at least a CSLG enzyme, wherein the genetically modified plant or plant part comprises elevated content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding non-genetically modified plant.

[00390] According to some embodiments, the polynucleotides disclosed herein are incorporated in a DNA construct enabling their expression in the plant cell. DNA constructs suitable for use in plants are known to a person skilled in the art. According to one embodiment, the DNA construct comprises at least one expression regulating element selected from the group consisting of a promoter, an enhancer, an origin of replication, a transcription termination sequence, a polyadenylation signal and the like.

[00391] The DNA constructs disclosed herein are designed according to the results to be achieved.

[00392] In crop plants, reduction of toxic steroidal alkaloids or steroidal glycoalkaloids or their

derivatives, metabolites, or biosynthetic intermediates thereof is desired in the edible parts of the plant, including for example, but not limited to fruits, seeds, roots, leaves, and tubers. On the other hand, enriching the content of toxic steroidal glycoalkaloids or biosynthetic intermediates in non-edible parts of the plant, including for example, but not limited to non-edible roots and leaves contributes to the resistance of the plant against a broad range of pathogens. In exemplary embodiments, plants overexpressing the steroidal glycoalkaloids can be used for producing them in the pharmaceutical industry.

[00393] In crop plants, modulation of steroidal saponins is also desired. A skilled artisan would appreciate that modulation of steroidal saponins may in certain embodiments, produce high-value products from a certain compound, for example but not limited to for food, cosmetcs and pharma industries. In some embodiments, modulation of steroidal saponins produces compounds in plants that will help in protection of plants against pathogens, for example but not limited to protection against bacetria, insect, virus, and insects. In some embodiments, modulation of steroidal saponins removes anti-nutritional compounds from plants.

[00394] In crop plants, reduction of toxic or bitter or hormone mimicking triterpenoid saponins or biosynthetic intermediates is desired in the edible parts of the plant, including, for example but not limited to, fruit and seeds. On the other hand, enriching the content of sweet tasting triterpenoid saponins or biosynthetic intermediates thereof is beneficial and could be used for producing natural sweeteners. Plants having increase content of triterpenoid saponins such as QS-21 can be used for producing adjuvants for the pharmaceutical industry.

[00395] According to yet additional embodiments, disclosed herein is a genetically modified or gene edited plant having enhanced expression of at least a CSLG gene, wherein the genetically modified or gene edited plant has an increased amount of at least one steroidal alkaloid, of at least one steroidal saponin, or of at least one triterpenoid saponin, compared to a corresponding unmodified or unedited plant. According to yet additional embodiments, disclosed herein is a genetically modified or gene edited plant having decreased expression of at least a CSLG gene, wherein the genetically modified or gene edited plant has a decreased amount of at least one steroidal alkaloid, a decreased amount of at least one steroidal saponin, a decreased amount of at least one triterpenoid saponin, compared to a corresponding unmodified or unedited plant.

[00396] The disclosure herein, and in the Examples below, show that by providing at least one heterologous gene encoding an enzyme or other protein involved in the biosynthetic pathway of a steroidal alkaloid, a steroidal saponin, and/or a triterpenoid saponin, for example but not limited to a heterologous CSLG, or by modifying an endogenous gene of the triterpenoid synthetic pathway, for example but not limited to a CSLG gene wherein said modification alters the expression thereof, or

the activity or stability of the encoded enzyme, the level of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof can be altered.

[00397] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00398] In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in a plant cell. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in an algal cell. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in a yeast. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in an insect cell. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in a bacterium. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in a plant. In some embodiments the amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal

saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is altered in a plant part.

[00399] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00400] For example, in some embodiments, as exemplified below in the Examples, a genetically modified cell, for example a yeast cell or a plant cell or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof having increased expression of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or a combination thereof, has been shown to have an increased level of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof. In other embodiments, as shown in the Examples, reducing or silencing expression of a saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG gene has been shown to result in a reduced level of the corresponding triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof.

[00401]

Steroidal Alkaloids, Derivatives Thereof, Metabolites Thereof, and Biosynthetic Intermediates Thereof

[00402] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, an *in vitro* biosynthetic system produces at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell as disclosed herein, comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprising at least one genetically modified cells comprises an increased content of at least one steroidal alkaloid, a derivative thereof,

a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof, and a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof.

[00403] In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell disclosed herein comprises producing an at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant or plant part comprises producing of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, disclosed herein is a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a plant or plant part comprising an at least one genetically modified cell as described herein. In some embodiments, disclosed herein is a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a a plant or plant part comprising an at least one genetically modified cell as described herein.

[00404] In the steroidal alkaloid biosynthetic pathway, steroidal alkaloids branch off from steroidal saponins after the formation of furostanol-type saponin aglycone (**FIGURE 1**). Further steps involving, at least in part, oxidation, transamination, ring formation, etc., result in biosynthesis of steroidal alkaloids (**FIGURE 1**).

[00405] Examples of steroidal glycoalkaloids include, but are not limited to, alpha-tomatine, tomatine, dehydrotomatine, alpha-chaonine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00406] In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterial. In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell. In some embodiments, the genetically modified cell comprising said increased content comprises a yeast cell. In some embodiments, the genetically modified cell comprising said increased content comprises an algal cell. In some embodiments, the genetically modified cell comprising said increased content comprises an insect cell. In some embodiments, the genetically modified cell comprising said increased content comprises a bacterium.

[00407] According to certain exemplary embodiments, the downstream steroidal glycoalkaloid is selected from the group consisting of esculeosides or dehydroesculeosides.

[00408] In some embodiments, a genetically modified plant disclosed herein comprises an

increased content of at least one steroidal alkaloid selected from any one of alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-tomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of tomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of dehydrotomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-chaconine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solanine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solasonine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solmargine.

[00409] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal alkaloid selected from any one of selected from any one of alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-tomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of tomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of dehydrotomatine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-chaconine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solanine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solasonine. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of alpha-solmargine.

[00410] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal and a decreased content of at least one steroidal alkaloid selected from any one of alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00411] In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid selected from any one of alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing alpha-tomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing tomatine. In some

embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing dehydrotomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing alpha-chaconine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing alpha-solanine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing alpha-solasonine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell comprises producing alpha-solmargine.

[00412] In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterium. In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in a plant cell. In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in a yeast cell. In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in an algal cell. In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in an insect cell. In some embodiments, the method of producing at least one steroidal alkaloid in a genetically modified cell comprises producing at least one steroidal alkaloid in a bacterium. In some embodiments, the plant cell is comprised in a plant or plant part.

[00413] In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing at least one steroidal alkaloid selected from any one of alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing alpha-tomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing tomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing dehydrotomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing alpha-chaconine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing alpha-solanine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing alpha-solasonine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing alpha-solmargine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing cholesterol. In some embodiments, a method of producing a steroidal alkaloid in a

genetically modified plant comprises producing 22-hydroxycholesterol. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing 22, 26-dihydroxycholesterol. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing furostanol-type saponin aglycone. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing furostanol-26-aldehyde. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing 26-amino-furostanol. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing tomatidenol. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing tomatidine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing tomatidine galactoside. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing gamma-tomatine. In some embodiments, a method of producing a steroidal alkaloid in a genetically modified plant comprises producing beta-1-tomatine.

[00414] In some embodiments, a method of producing a steroidal alkaloid in an *in vitro* translation system comprises producing at least one steroidal alkaloid selected from any one of alpha-tomatine, dehydrotomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing alpha-tomatine. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing tomatine. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing alpha-chaconine. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing alpha-solanine. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing alpha-solasonine. In some embodiments, a method of producing a steroidal alkaloid in an in vitro translation system comprises producing alpha-solmargine.

[00415] In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least one steroidal alkaloid selected from any one of alpha-tomatine, dehydrotomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a method of reducing the content of at least one steroidal alkaloid comprising alpha-tomatine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing tomatine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing dehydrotomatine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing alpha-chaconine. In some embodiments, a method

of reducing at least one steroidal alkaloid comprises reducing alpha-solanine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing alpha-solasonine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing alpha-solmargine.

[00416] In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least one steroidal alkaloid selected from any one of of alpha-tomatine, dehydrotomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a method of increasing the content of at least one steroidal alkaloid comprising medicagenic acid alpha-tomatine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing tomatine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing alpha-chaconine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing alpha-solanine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing alpha-solasonine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing alpha-solmargine.

[00417] In some embodiments, the content of both at least one steroidal alkaloid and at least one steroidal alkaloid biosynthetic intermediate are altered. In some embodiments, an at least one steroidal alkaloid and an at least one steroidal alkaloid biosynthetic intermediate are increased. In some embodiments, an at least one steroidal alkaloids and an at least one steroidal alkaloid biosynthetic intermediate are decreased. In some embodiments, an at least one steroidal alkaloid is increased and an at least one steroidal alkaloid biosynthetic intermediate is decreased. In some embodiments, an at least one steroidal alkaloid is decreased and an at least one steroidal alkaloid biosynthetic intermediate is increased. In some embodiments, the content of a steroidal alkaloid is altered without measurably altering the content of a steroidal alkaloid intermediate. In some embodiments, the content of a steroidal alkaloid intermediate is altered without measurably altering the content of a steroidal alkaloid.

[00418] The term "intermediate" may be used interchangeably in some embodiments with the term "biosynthetic intermediate", having all the same qualities and meanings.

[00419] In some embodiments, a biosynthetic intermediate of a steroidal alkaloid comprises cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof (**FIGURE 1**).

[00420] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising cholesterol, 22-

hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising cholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising 22-hydroxycholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising22,26-dihydroxycholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprisingfurostanol-type saponin aglycone. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising furostonal-26-aldehyde. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising 26-amino-furostanol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising tomatidenol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising tomatidine. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising tomatidine galactoside. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising gamma-tomatine. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising beta-1-tomatine.

[00421] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof . In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising cholesterol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising 22-hydroxysholesterol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising22,26-dihydroxycholesterol. In some embodiments, a genetically modified plant comprises a decreased

content of at least one biosynthetic intermediate of a steroidal alkaloid comprisingfurostanol-type saponin aglycone. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising furostanol-26-aldehyde. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising26-amino-furstanol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprisingtomatidenol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprisingtomatidine. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising tomatidine galactoside. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising gamma-tomatine. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal alkaloid comprising beta-1-tomatine.

[00422] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid biosynthetic intermediate and a decreased content of at least one steroidal alkaloid biosynthetic intermediate, said intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof.

[00423] In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least one steroidal alkaloid biosynthetic intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof. . In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least cholesterol. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least 22-hydroxycholesterol. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least 22,26-dihydroxycholesterol.In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least furostanol-type saponin aglycone. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least furostonal-26-aldehyde. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least 26-amino-furostanol. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least tomatidenol. In some embodiments, a method of reducing at least one

steroidal alkaloid comprises reducing at tomatidine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least tomatidine galactoside. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at least gamma-tomatine. In some embodiments, a method of reducing at least one steroidal alkaloid comprises reducing at leastbeta-1-tomaine.

[00424] In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least one steroidal alkaloid biosynthetic intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least cholesterol. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least 22-hydroxycholesterol. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least22,26-dihydroxycholesterol. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least furostanol-type saponin aglycone.In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least furostanol-26-aldehyde. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least 26-amino-furostanol. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least tomatidenol. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least tomatidine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least tomatidine galactoside. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least gamma-tomatine. In some embodiments, a method of increasing at least one steroidal alkaloid comprises increasing at least beta-1-tomatine.

[00425] In some embodiments, steroidal alkaloids biosynthetic intermediate comprises any cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or a combination thereof .

[00426] Unexpectedly, the present disclosure now shows that levels of steroidal alkaloids or their derivatives, metabolites, or biosynthetic intermediates can be increased in cells, for example plant cells or yeast cells or algal cells, or insect cells, or bacterium, or plants by genetically modifying the cell or plant to express at least one heterologous gene encoding an enzyme, for example an enzyme or enzymes of the steroidal alkaloid biosynthetic pathway. In other embodiments, described and exemplified herein are methods of genetically modifying an at least one endogenous gene in a plant

cell, for example but not limited to a CSLG gene, to regulate expression, activity, or stability, or any combination thereof. The Examples below disclose enzyme activities and enzymes previously unknown to be part of the steroidal alkaloid biosynthetic pathway. Without this knowledge, production of the steroidal alkaloid compounds was not possible.

[00427] In some embodiments, the steroidal alkaloid metabolic pathway can result in cells or plants comprising elevated content of steroidal alkaloids, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or in plants having an increased content of these compounds in the plant or plant parts. In some embodiments, plants including but not limited to crop plants are produced, wherein the crop has an increased content of a useful steroidal alkaloid or steroidal alkaloids. In some embodiments, disclosed herein are the means and methods for producing cells including but not limited to plant cells, yeast, or algal cells; plants including but not limited to crop plants, or a part of a plant; having increased levels of a steroidal alkaloid, or steroidal alkaloids, or derivatives thereof, or metabolites thereof, or a biosynthetic intermediate thereof. Alternatively, or additionally, controlling the expression of genes disclosed herein may be used for the production of desired steroidal alkaloids for further use, for example in the pharmaceutical industry or for the formulation of dietary or other supplements, for example but not limited to sweeteners. In some embodiments, these high value saponins may be purified and used, e.g., as sweeteners, foaming agents, emulsifiers, preservatives, anti-carcinogens, hypocholesterolemic agents, anti-inflammatory agents, anti-oxidants, biological adjuvants, anti-microbial agents, insecticidal agents, anti-feedants, or anti-fungal agents, or any combination thereof. The cells and plants disclosed herein comprise compounds of significant nutritional, pharmaceutical, and commercial value.

[00428] In some embodiments, derivatives of steroidal alkaloids comprise glycosylated derivatives of steroidal alkaloids.

[00429] In some embodiments, a genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, as described herein, comprises an altered content of at least a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell or unmodified plant. In some embodiments, an altered content comprises an increased content. In some embodiments, for example, the genetically modified cell or genetically modified plant has an increased content of at least a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof.

[00430] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least one steroidal alkaloid. In some embodiments, the genetically modified cell or genetically modified plant

or plant part comprising at least one genetically modified plant cell, has an increased content of at least two steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six steroidal alkaloids.

[00431] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a derivative of a steroidal alkaloid. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two derivatives of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three derivatives of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four derivatives of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five derivatives of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six derivatives of a steroidal alkaloid or of steroidal alkaloids.

[00432] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a metabolite of a steroidal alkaloid. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two metabolites of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three metabolites of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or

genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four metabolites of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five metabolites of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six metabolites of a steroidal alkaloid or of steroidal alkaloids.

[00433] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a biosynthetic intermediate of a steroidal alkaloid. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids.

[00434] As skilled artisan would recognize that the terms “content” and “level” in reference to a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, encompasses the quantity of the compound, for example the quantity of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof in a genetically modified cell or in a genetically modified plant, compared with a control cell or control plant. In this context, the terms “content” and “level” may be used interchangeably having all the same meanings and qualities.

[00435] In some embodiments, a steroidal alkaloid having increased content comprises a nutrition, cosmetic, or pharmaceutical agent, or any combination thereof. In some embodiments, a steroidal alkaloid having increased content comprises steroidal alkaloid selected from an alpha-tomatine,

tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a steroidal alkaloid having increased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00436] In some embodiments, a steroidal alkaloid having increased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00437] In some embodiments, a steroidal alkaloid having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00438] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of alpha-tomatine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of tomatine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of dehydrotomatine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of alpha-chaconine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of alpha-solanine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of alpha-solasonine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of alpha-solmargine. A skilled artisan would appreciate that in some embodiments, instances wherein the content of one steroidal alkaloid is increased additional steroidal alkaloids, or intermediates, or a combination thereof may also be increased in the same time cell.

[00439] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or

an algal cell, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of at least two biosynthetic intermediates of a steroidal alkaloid or of steroidal alkaloids, said intermediates comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof.

[00440] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising cholesterol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising 22-hydroxycholesteerol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising 22,26-dihydroxysholesterol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising furostanol-type saponin aglycone. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising furostanol-26-aldehyde. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising 26-amino-furostanol, or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising

tomatidenol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkyloid, said intermediate comprising tomatidine. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising tomatidine galactoside, or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising gamma-tomatine, or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising beta-1-tomatine, or any combination thereof.

[00441] In some embodiments, a steroidal alkaloid having decreased content in a plant or plant part comprising at least one genetically modified cell, comprises a compound having a bitter taste or a toxin. In some embodiments, a steroidal alkaloid having decreased content comprises a steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, a steroidal alkaloid having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloids selected from In some embodiments, a steroidal saponin having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises a steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00442] In some embodiments, a steroidal alkaloid having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloids selected from In some embodiments, a steroidal alkaloid having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises a steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00443] In some embodiments, a steroidal alkaloid having decreased content in a genetically modified plant or part thereof comprising at least one genetically modified cell, comprises a steroidal alkaloid selected from alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine,

alpha-solasonine, alpha-solmargine, or any combination thereof.

[00444] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of alpha-tomatine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of tomatine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of dehydrotomatine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of alpha-chaconine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of alpha-solanine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of alpha-solasonine.

[00445] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, beta-1-tomatine, or any combination thereof.

[00446] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising cholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising 22-hydroxycholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising 22,26-dihydroxycholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising fursotanol-type saponin aglycone. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising furostanol-26-aldehyde. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising

26-amino-furostanol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising tomatidenol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising tomatidine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising tomatidine galactoside. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising gamma-tomatine. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal alkaloid, said intermediate comprising beta-1-tomatine.

[00447] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00448] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00449] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite

thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00450] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00451] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00452] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a decreased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00453] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite

thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00454] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00455] In some embodiments, a method of producing at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00456] In some embodiments, a method of producing at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00457] In some embodiments, a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal alkaloid, a derivative

thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00458] In some embodiments, a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00459] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an altered content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00460] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a

phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

Steroidal Saponins, Derivatives Thereof, Metabolites Thereof, and Biosynthetic Intermediates Thereof

[00461] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, an *in vitro* biosynthetic system produces at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell as disclosed herein, comprises a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprising at least one genetically modified cells comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof, and a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof.

[00462] In some embodiments, a method of producing a steroidal alkaloid in a genetically modified cell disclosed herein comprises producing an at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a method of producing a steroidal saponin in a genetically modified plant or plant part comprises producing of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, disclosed herein is a method of reducing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a plant or

plant part comprising an at least one genetically modified cell as described herein. In some embodiments, disclosed herein is a method of increasing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a a plant or plant part comprising an at least one genetically modified cell as described herein.

[00463] Steroidal saponins branch off from steroidal alkaloids after the formation of furostanol-type saponin aglycone (**FIGURE 1**).

[00464] The commonly used nomenclature for saponins distinguishes between triterpenoid saponins (also called triterpene saponins) and steroidal saponins, which is based on the structure and biochemical background of their aglycones. Both sapogenin types are thought to derive from 2,3-oxidosqualene, a central metabolite in sterol biosynthesis. In phytosterol anabolism, 2,3-oxidosqualene is mainly cyclized into cycloartenol. "Steroidal sapogenins" are thought to derive from intermediates in the phytosterol pathway downstream of cycloartenol formation.

[00465] Steroidal saponins, include, but are not limited to, uttroside B, a tomatoside, or any combination thereof.

[00466] In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterial. In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell. In some embodiments, the genetically modified cell comprising said increased content comprises a yeast cell. In some embodiments, the genetically modified cell comprising said increased content comprises an algal cell. In some embodiments, the genetically modified cell comprising said increased content comprises an insect cell. In some embodiments, the genetically modified cell comprising said increased content comprises a bacterium.

[00467] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal saponin selected from any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of uttroside B.

[00468] In some embodiments, derivatives of steroidal saponins comprise glycosylated derivatives of steroidal saponins.

[00469] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal saponin selected from any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of uttroside B.

[00470] In some embodiments, a genetically modified plant disclosed herein comprises an

increased content of at least one steroidal saponin and a decreased content of at least one steroidal saponin selected from any one of uttroside B, a tomatoside, or any combination thereof.

[00471] In some embodiments, a method of producing a steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponin selected from any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a method of producing a steroidal saponin in a genetically modified cell comprises producing uttroside B.

[00472] In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterium. In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in a plant cell. In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in a yeast cell. In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in an algal cell. In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in an insect cell. In some embodiments, the method of producing at least one steroidal saponin in a genetically modified cell comprises producing at least one steroidal saponins in a bacterium. In some embodiments, the plant cell is comprised in a plant or plant part.

[00473] In some embodiments, a method of producing a steroidal saponin in a genetically modified plant comprises producing at least one steroidal saponin selected from any one of any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a method of producing a steroidal saponin in a genetically modified plant comprises producing uttroside B.

[00474] In some embodiments, a method of producing a steroidal saponin in an *in vitro* translation system comprises producing at least one steroidal saponin selected from any one of any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a method of producing a steroidal saponin in an in vitro translation system comprises producing uttroside B.

[00475] In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least one steroidal saponin selected from any one of any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, a method of reducing the content of at least one steroidal saponin comprising uttroside B.

[00476] In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least one steroidal saponin selected from any one of any one of uttroside B, a tomatoside, or any combination thereof. In some embodiments, the at least one steroidal saponin comprises uttroside B. In some embodiments, the at least one steroidal saponin comprises a tomatoside.

[00477] In some embodiments, the content of both at least one steroidal saponin and at least one steroidal saponin biosynthetic intermediate are altered. In some embodiments, an at least one steroidal saponin and an at least one steroidal saponin biosynthetic intermediate are increased. In some embodiments, an at least one steroidal saponins and an at least one steroidal saponin biosynthetic intermediate are decreased. In some embodiments, an at least one steroidal saponin is increased and an at least one steroidal saponin biosynthetic intermediate is decreased. In some embodiments, an at least one steroidal saponin is decreased and an at least one steroidal saponin biosynthetic intermediate is increased. In some embodiments, the content of a steroidal saponin is altered without measurably altering the content of a steroidal saponin intermediate. In some embodiments, the content of a steroidal saponin intermediate is altered without measurably altering the content of a steroidal saponin.

[00478] The term "intermediate" may be used interchangeably in some embodiments with the term "biosynthetic intermediate", having all the same qualities and meanings.

[00479] In some embodiments, a biosynthetic intermediate of a steroidal saponin comprises cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof.

[00480] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one biosynthetic intermediate of a steroidal saponin comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal saponin comprising cholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal saponin comprising 22-hydroxycholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal saponin comprising22,26-dihydroxycholesterol. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a steroidal saponin comprisingfurostanol-type-saponin aglycone.

[00481] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one biosynthetic intermediate of a steroidal saponin comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal saponin comprising cholesterol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal saponin comprising 22-hydroxycholesterol. In some embodiments, a

genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal saponin comprising 22,26-dihydroxycholeseterol. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a steroidal saponin comprising furostanol-type saponin aglycone.

[00482] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal saponin biosynthetic intermediate and a decreased content of at least one steroidal saponin biosynthetic intermediate, said intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof.

[00483] In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least one steroidal saponin biosynthetic intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof. In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least cholesterol. In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least 22-hydroxycholesterol. In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least 22,26-dihydroxycholesterol. In some embodiments, a method of reducing at least one steroidal saponin comprises reducing at least furostanol-type saponin aglycone.

[00484] In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least one steroidal saponin biosynthetic intermediate selected from any one of cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or any combination thereof. In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least cholesterol. In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least 22-hydroxycholesterol. In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least 22,26-dihydroxycholesterol. In some embodiments, a method of increasing at least one steroidal saponin comprises increasing at least furostanol-type saponin aglycone.

[00485] In some embodiments, steroidal saponins biosynthetic intermediate comprises any cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, or a combination thereof .

[00486] Unexpectedly, the present disclosure now shows that levels of steroidal saponins or their derivatives, metabolites, or biosynthetic intermediates can be increased in cells, for example plant cells or yeast cells or algal cells, or insect cells, or bacterium, or plants by genetically modifying the cell or plant to express at least one heterologous gene encoding an enzyme, for example an enzyme

or enzymes of the steroidal saponin biosynthetic pathway. In other embodiments, described and exemplified herein are methods of genetically modifying an at least one endogenous gene in a plant cell, for example but not limited to a CSLG gene, to regulate expression, activity, or stability, or any combination thereof. The Examples below disclose enzyme activities and enzymes previously unknown to be part of the steroidal saponin biosynthetic pathway. Without this knowledge, production of the steroidal saponin compounds was not possible.

[00487] In some embodiments, the steroidal saponin metabolic pathway can result in cells or plants comprising elevated content of steroidal saponins, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or in plants having an increased content of these compounds in the plant or plant parts. In some embodiments, plants including but not limited to crop plants are produced, wherein the crop has an increased content of a useful steroidal saponin or steroidal saponins. In some embodiments, disclosed herein are the means and methods for producing cells including but not limited to plant cells, yeast, or algal cells; plants including but not limited to crop plants, or a part of a plant; having increased levels of a steroidal saponin, or steroidal saponins, or derivatives thereof, or metabolites thereof, or a biosynthetic intermediate thereof. Alternatively, or additionally, controlling the expression of genes disclosed herein may be used for the production of desired steroidal saponins for further use, for example in the pharmaceutical industry or for the formulation of dietary or other supplements, for example but not limited to sweeteners. In some embodiments, these high value saponins may be purified and used, e.g., as sweeteners, foaming agents, emulsifiers, preservatives, anti-carcinogens, hypocholesterolemic agents, anti-inflammatory agents, anti-oxidants, biological adjuvants, anti-microbial agents, insecticidal agents, anti-feedants, or anti-fungal agents, or any combination thereof. The cells and plants disclosed herein comprise compounds of significant nutritional, pharmaceutical, and commercial value.

[00488] In some embodiments, a genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, as described herein, comprises an altered content of at least a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell or unmodified plant. In some embodiments, an altered content comprises an increased content. In some embodiments, for example, the genetically modified cell or genetically modified plant has an increased content of at least a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof.

[00489] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least one steroidal saponin. In some embodiments, the genetically modified cell or genetically modified plant

or plant part comprising at least one genetically modified plant cell, has an increased content of at least two steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six steroidal saponins.

[00490] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a derivative of a steroidal saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two derivatives of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three derivatives of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four derivatives of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five derivatives of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six derivatives of a steroidal saponin or of steroidal saponins.

[00491] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a metabolite of a steroidal saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two metabolites of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three metabolites of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or

genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four metabolites of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five metabolites of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six metabolites of a steroidal saponin or of steroidal saponins.

[00492] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a biosynthetic intermediate of a steroidal saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two biosynthetic intermediates of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three biosynthetic intermediates of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four biosynthetic intermediates of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five biosynthetic intermediates of a steroidal saponin or of steroidal saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six biosynthetic intermediates of a steroidal saponin or of steroidal saponins.

[00493] As skilled artisan would recognize that the terms “content” and “level” in reference to a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, encompasses the quantity of the compound, for example the quantity of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof in a genetically modified cell or in a genetically modified plant, compared with a control cell or control plant. In this context, the terms “content” and “level” may be used interchangeably having all the same meanings and qualities.

[00494] In some embodiments, a steroidal saponin having increased content comprises a nutrition, cosmetic, or pharmaceutical agent, or any combination thereof. In some embodiments, a steroidal saponin having increased content comprises steroidal saponin selected from an uttroside B, a

tomatosideor any combination thereof. In some embodiments, a steroidal alkaloid having increased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloid selected from uttroside B, a tomatoside, or any combination thereof.

[00495] In some embodiments, a steroidal saponin having increased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal alkaloid selected from uttroside B, a tomatosideor any combination thereof.

[00496] In some embodiments, a steroidal saponin having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises steroidal saponin selected from uttroside B, a tomatosideor any combination thereof.

[00497] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of uttroside B. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of a tomatoside.

[00498] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal steroidal saponin, said intermediate comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of at least two biosynthetic intermediates of a steroidal saponin or of steroidal saponins, said intermediates comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, or any combination thereof.

[00499] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising cholesterol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising 22-hydroxycholesteerol. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising 22,26-dihydroxysholesterol. In some embodiments,

a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising furostanol-type saponin aglycone.

[00500] In some embodiments, a steroidal saponin having decreased content in a plant or plant part comprising at least one genetically modified cell, comprises a compound having a bitter taste or a toxin. In some embodiments, a steroidal saponin having decreased content comprises a steroidal saponin selected from uttroside B, a tomatosideor any combination thereof. In some embodiments, a steroidal saponin having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal saponins selected from In some embodiments, a steroidal saponin having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises steroidal saponin selected from uttroside B, a tomatosideor any combination thereof.

[00501] In some embodiments, a steroidal saponin having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more steroidal saponins selected from In some embodiments, a steroidal saponin having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises a steroidal saponin selected from uttroside B, a tomatosideor any combination thereof.

[00502] In some embodiments, a steroidal saponin having decreased content in a genetically modified plant or part thereof comprising at least one genetically modified cell, comprises a steroidal saponin selected from uttroside B, a tomatosideor any combination thereof.

[00503] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of uttroside B. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a tomatoside.

[00504] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type-saponin aglycone, or any combination thereof.

[00505] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising cholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising

22-hydroxycholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising 22, 26-dihydroxycholesterol. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a steroidal saponin, said intermediate comprising furostanol-type saponin aglycone.

[00506] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00507] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00508] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00509] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite

thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00510] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00511] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a decreased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00512] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00513] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite

thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00514] In some embodiments, a method of producing at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00515] In some embodiments, a method of producing at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00516] In some embodiments, a method of reducing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00517] In some embodiments, a method of reducing the content of at least one steroidal saponin,

a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00518] In some embodiments, a method of increasing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an altered content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00519] In some embodiments, a method of increasing the content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

Triterpenoid Saponins, Derivatives Thereof, Metabolites Thereof, and Biosynthetic Intermediates Thereof

[00520] In some embodiments, a genetically modified cell disclosed herein comprises an increased

content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell disclosed herein comprises an increased content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, an *in vitro* biosynthetic system produces at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant comprising at least one genetically modified cell as disclosed herein, comprises a decreased content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprising at least one genetically modified cells comprises an increased content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof, and a decreased content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof.

[00521] In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell disclosed herein comprises producing an at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant or plant part comprises producing of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof. In some embodiments, disclosed herein is a method of reducing the content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a plant or plant part comprising an at least one genetically modified cell as described herein. In some embodiments, disclosed herein is a method of increasing the content of at least one triterpenoids saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, of any combination thereof in a plant or plant part comprising an at least one genetically modified cell as described herein.

[00522] The commonly used nomenclature for saponins distinguishes between triterpenoid saponins (also called triterpene saponins) and steroidal saponins, which is based on the structure and biochemical background of their aglycones. Both sapogenin types are thought to derive from 2,3-

oxidosqualene, a central metabolite in sterol biosynthesis. In phytosterol anabolism, 2,3-oxidosqualene is mainly cyclized into cycloartenol. "Triterpenoid sapogenins" branch off the phytosterol pathway by alternative cyclization of 2,3-oxidosqualene, while "steroidal sapogenins" are thought to derive from intermediates in the phytosterol pathway downstream of cycloartenol formation (see **FIGURE 3A**).

[00523] Saponins may be selected from the group comprising dammarane-type saponins, tirucallane-type saponins, lupane-type saponins, oleanane-type saponins, taraxasterane-type saponins, ursane-type saponins, hopane-type saponins, cucurbitane-type saponins, cycloartane-type saponins, lanostane-type saponins, and steroid-type saponins. The aglycon backbones, the sapogenins, can be similarly classified and may be selected from the group comprising dammarane-type sapogenins, tirucallane-type sapogenins, lupane-type sapogenins, oleanane-type sapogenins, taraxasterane-type sapogenins, ursane-type sapogenins, hopane-type sapogenins, cucurbitane-type sapogenins, cycloartane-type sapogenins, lanostane-type sapogenins, and steroid-type sapogenins.

[00524] In some embodiments, the "triterpenoid sapogenins" disclosed herein may be selected from the group comprising dammarane-type sapogenins, tirucallane-type sapogenins, lupane-type sapogenins, oleanane-type sapogenins, ursane-type sapogenins, and hopane-type sapogenins. In some embodiment, the triterpenoid sapogenins as produced by the method of the disclosure are dammarane-type sapogenins, or tirucallane-type sapogenins, or lupane-type sapogenins, or oleanane-type sapogenins, or ursane-type sapogenins, or hopane-type sapogenins.

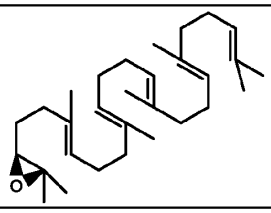
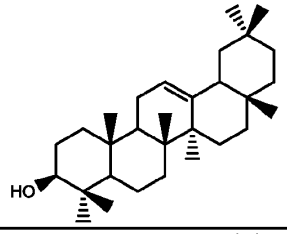
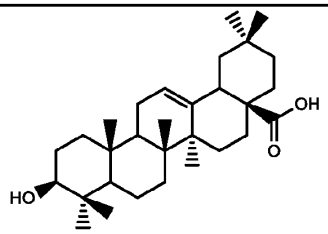
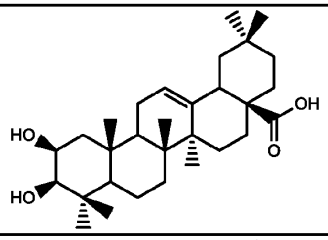
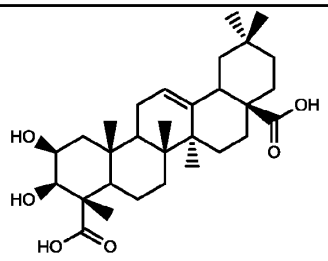
[00525] Triterpenoid sapogenins typically have a tetracyclic or pentacyclic skeleton. In some embodiments, the sapogenin building blocks themselves may have multiple modifications, for example but not limited to small functional groups, including hydroxyl, keto, aldehyde, and carboxyl moieties, of precursor sapogenin backbones such as βamyrin, lupeol, and dammarenediol.

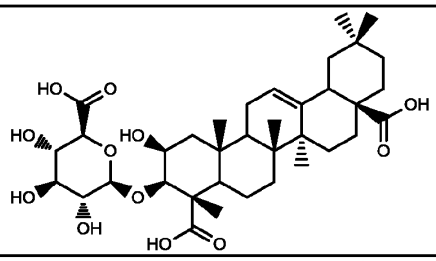
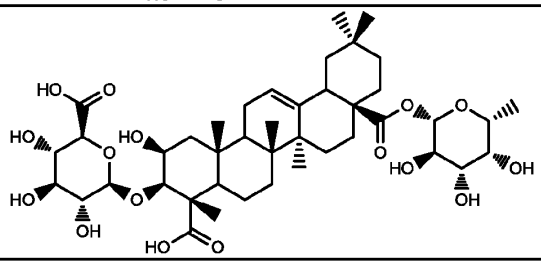
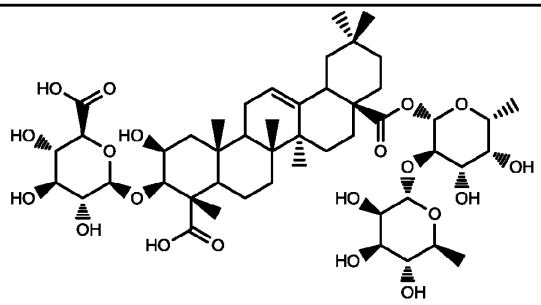
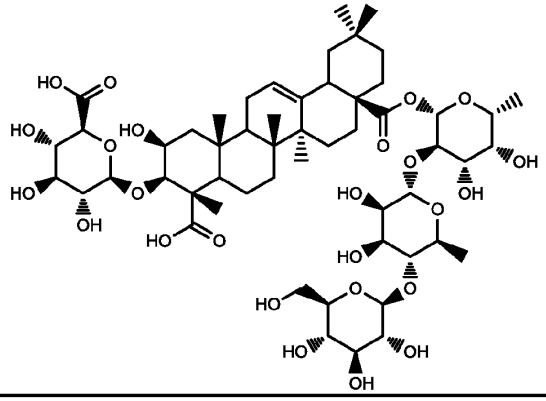
[00526] The terms "triterpene" and "triterpenoid" may be used interchangeably having the same meaning and qualities. Triterpenoid saponins comprise multiple functions and may be used in different roles including but not limited to as a sweetener, a foaming agent, an emulsifier, a preservative, an anti-carcinogen, a hypocholesterolemic agent, an anti-inflammatory agent, an anti-oxidant, a biological adjuvant, an anti-microbial agent, an insecticidal agent, an antifeedant, an anti-fungal agent, or any combination thereof.

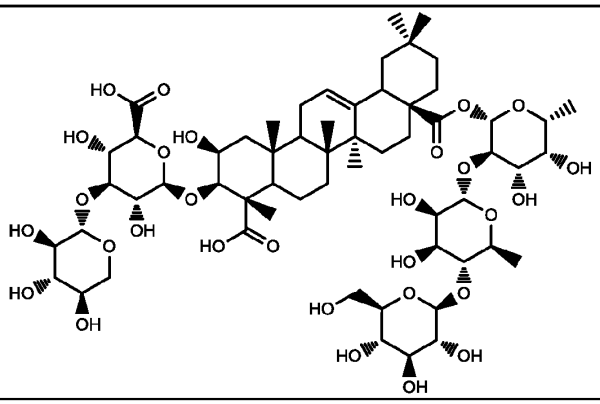
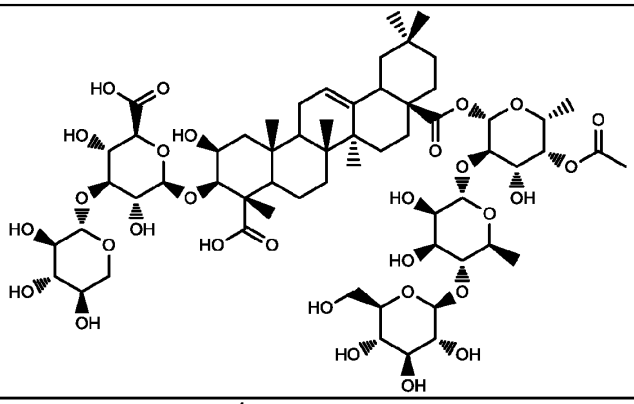
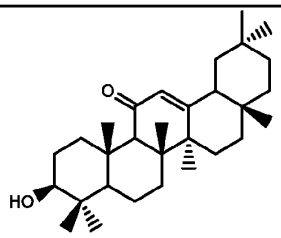
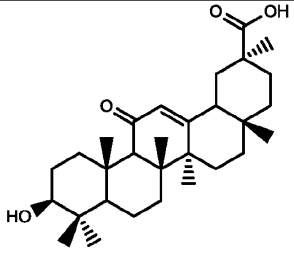
[00527] In some embodiments, a triterpenoid sapogenins, as disclosed herein, also encompass new-to-nature triterpenoid compounds, which are structurally related to the naturally occurring triterpenoid sapogenins. These new-to-nature triterpenoid sapogenins may be novel compounds that can be obtained after genetic engineering of the synthesizing eukaryotic host cell, for example a plant cell or a yeast cell.

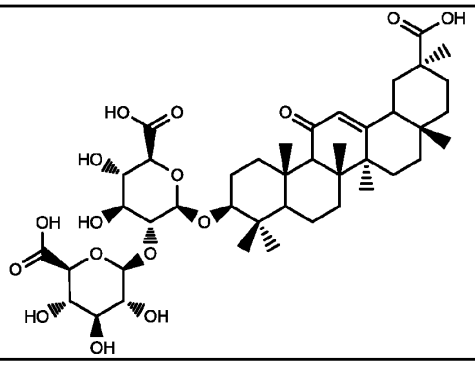
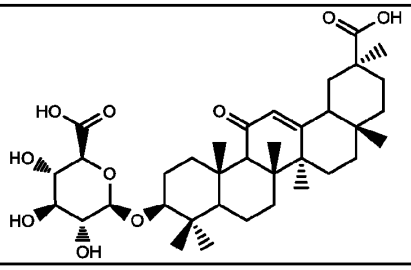
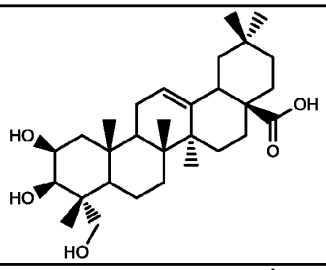
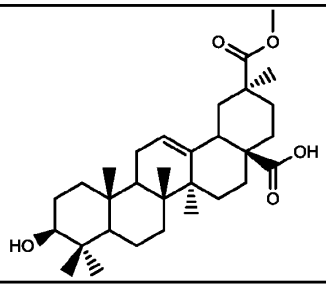
[00528] Examples of triterpenoid saponins and biosynthetic intermediates thereof are presented in **Table 2** below.

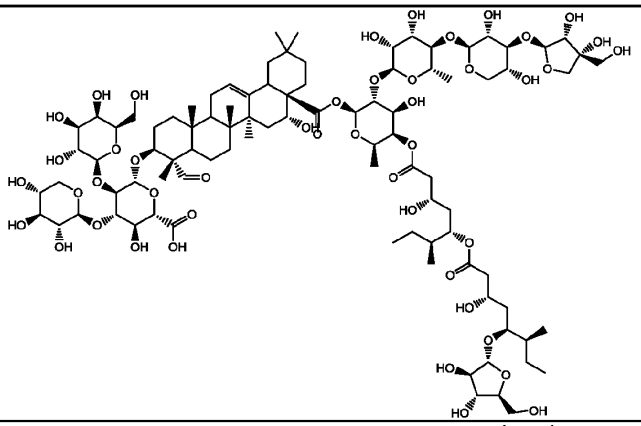
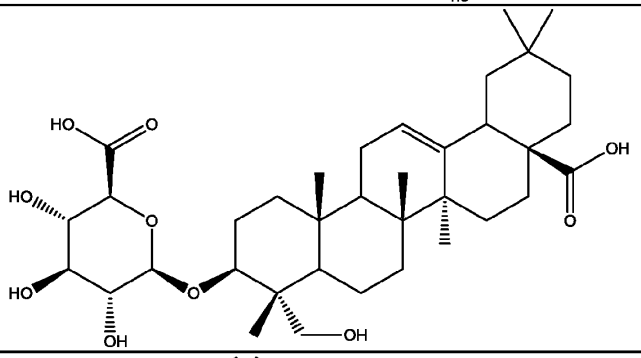
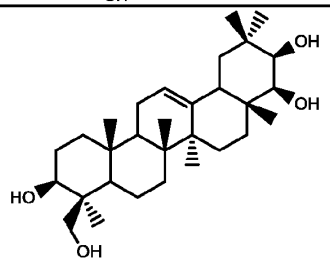
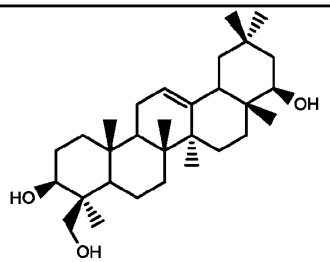
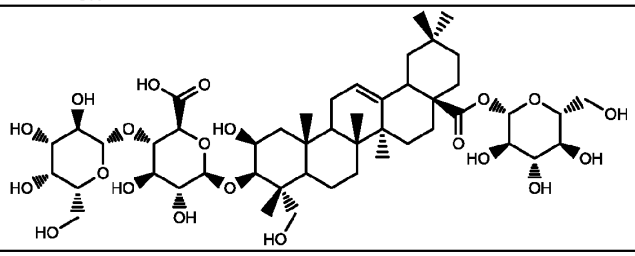
[00529] **Table 2: Triterpenoid Saponins and Biosynthetic Intermediates Thereof**

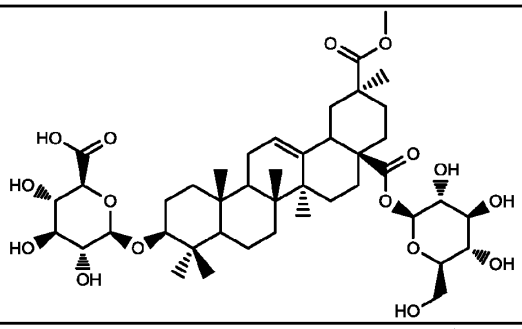
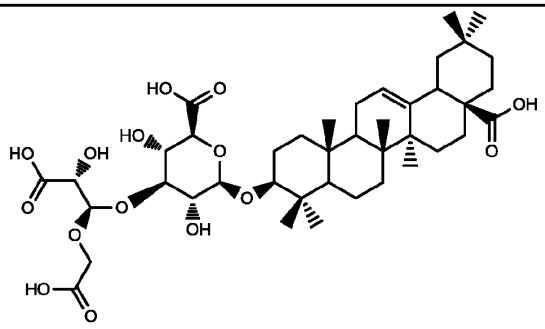
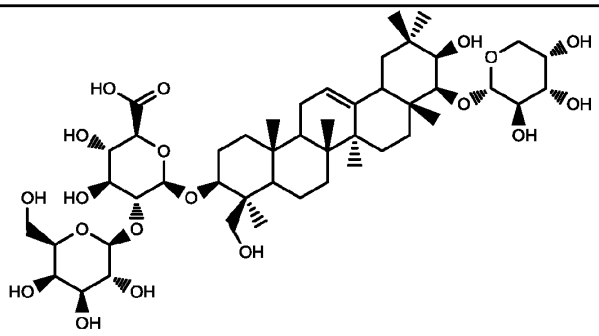
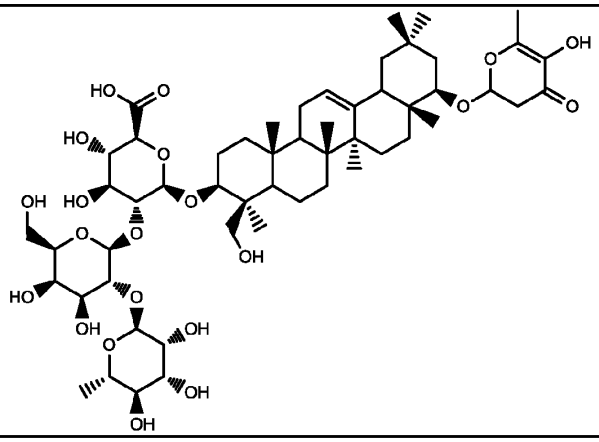
COMPOUND	STRUCTURE	NAME
1		(S)-2,3-Oxidosqualene
2		β-Amyrin
3		Oleanolic acid (OA)
4		Augustic acid (AA)
5		Medicagenic acid (MA)

COMPOUND	STRUCTURE	NAME
6		Medicagenic acid 3-glucuronide
7		Yossoside I (3-O-β-D-glucuronopyranosyl-28-O-β-D-fucopyranosyl-medicagenic acid)
8		Yossoside II (3-O-[β-D-glucuronopyranosyl]-28-O-[α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid)
9		Yossoside III (3-O-[β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid)

COMPOUND	STRUCTURE	NAME
10		Yossoside IV (3-O-[β-D-xylopyranosyl-(1–>3)-β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-β-D-fucopyranosyl]-medicagenic acid)
11		Yossoside V (3-O-[β-D-xylopyranosyl-(1–>3)-β-D-glucuronopyranosyl]-28-O-[β-D-glucopyranosyl-(1–>4)-α-L-rhamnopyranosyl-(1–>2)-4-acetyl-β-D-fucopyranosyl]-medicagenic acid)
12		11-oxo-β-Amyrin
13		glycyrrhetinic acid

COMPOUND	STRUCTURE	NAME
14		glycyrrhizin
15		glycyrrhetinic acid 3-O-monoglucuronide
16	not available	Yossoside Va
17	not available	Yossoside VI
18	not available	Yossoside VII
19	not available	Yossoside VIIa
20	not available	Yossoside VIII
21	not available	Yossoside IX
22	not available	Yossoside X
23	not available	Yossoside XI
24	not available	Yossoside XII
25		Bayogenin
26		Serjanic acid

COMPOUND	STRUCTURE	NAME
27		QS-21
28		Hederagenin-3GlcA
29		Soyasapogenol A
30		Soyasapogenol B
31		Bayogenin-hexA-hex-hex

COMPOUND	STRUCTURE	NAME
32		Serjanic acid GlcA-glc
33		betavulgaroside IV
34		Soyasapogenol A hexA-hex-pent
35		soyasaponin VI

[00530] In some embodiments, derivatives of triterpenoid saponins comprise glycosylated

derivatives of triterpenoid saponins.

[00531] In some embodiments, a triterpenoid saponins disclosed herein comprise any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), glycyrrhetinic acid 3-O-monoglucuronide (Compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. .

[00532] In some embodiments, a triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a triterpenoid saponin comprises Compound 7. In some embodiments, a triterpenoid saponin comprises Compound 8. In some embodiments, a triterpenoid saponin comprises Compound 9. In some embodiments, a triterpenoid saponin comprises Compound 10. In some embodiments, a triterpenoid saponin comprises Compound 11. In some embodiments, a triterpenoid saponin comprises glycyrrhizin (Compound 14). In some embodiments, a triterpenoid saponin comprises glycyrrhetinic acid 3-O-monoglucuronide (Compound 15). In some embodiments, a triterpenoid saponin comprises bayogenin (Compound 25). In some embodiments, a triterpenoid saponin comprises bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a triterpenoid saponin comprises serjanic acid (Compound 26). In some embodiments, a triterpenoid saponin comprises serjanic acid-hexA-hex (Compound 32). In some embodiments, a triterpenoid saponin comprises soyasapogenol A (Compound 29). In some embodiments, a triterpenoid saponin comprises soyasapogenol B (Compound 30). In some embodiments, a triterpenoid saponin comprises soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a triterpenoid saponin comprises soyasaponin VI (Compound 35). In some embodiments, a triterpenoid saponin comprises betavulgaroside IV (Compound 33). In some embodiments, a triterpenoid saponin comprises hederagenin-3GlcA. In some embodiments, a triterpenoid saponin comprises gypsogenin-3GlcA. In some embodiments, a triterpenoid saponin comprises gypsogenic acid-3GlcA. In some embodiments, a triterpenoid saponin comprises a QS-21 adjuvant.

[00533] Altering the content of a triterpenoid saponin may in some embodiments, be commercially beneficial, for example but not limited to the ability to produce the sweetener, glycyrrhizin (Compound 14). In some embodiments, it is commercially beneficial to produce a triterpenoid saponin comprising any of a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen,

hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof. In other embodiments, altering the content of a triterpenoid saponin may be beneficial, for example but not limited to reducing or eliminating bitter tasting triterpenoid saponins, for example but not limited to in quinoa, which is produced from a *Chenopodium quinoa* plant in the *Caryophyllales* order of plants. In other embodiments, altering the content of a triterpenoid saponin may be beneficial, for example but not limited to reducing or eliminating triterpenoid saponins that have hormone mimicking properties.

[00534] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Compound 7. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Compound 8. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Compound 9. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Compound 10. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Compound 11. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of glycyrrhizin (Compound 14). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of bayogenin (Compound 25). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of serjanic acid (Compound 26). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of serjanic acid-hexA-hex (Compound 32). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of soyasapogenol A

(Compound 29). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of soyasapogenol B (Compound 30). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of soyasaponin VI (Compound 35). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of betavulgaroside IV (Compound 33). In some embodiments, a genetically modified cell disclosed herein comprises an increased content of hederagenin-3GlcA. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of gypsogenin-3GlcA. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of gypsogenic acid-3GlcA. In some embodiments, a genetically modified cell disclosed herein comprises an increased content of a QS-21 adjuvant.

[00535] In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterial. In some embodiments, the genetically modified cell comprising said increased content comprises a plant cell. In some embodiments, the genetically modified cell comprising said increased content comprises a yeast cell. In some embodiments, the genetically modified cell comprising said increased content comprises an algal cell. In some embodiments, the genetically modified cell comprising said increased content comprises an insect cell. In some embodiments, the genetically modified cell comprising said increased content comprises a bacterium.

[00536] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of Compound 7. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of Compound 8. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of Compound 9. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of

Compound 10. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of Compound 11. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of glycyrrhizin (Compound 14). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of bayogenin (Compound 25). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of serjanic acid (Compound 26). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of serjanic acid-hexA-hex (Compound 32). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of soyasapogenol A (Compound 29). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of soyasapogenol B (Compound 30). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of soyasaponin VI (Compound 35). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of betavulgaroside IV (Compound 33). In some embodiments, a genetically modified plant disclosed herein comprises an increased content of hederagenin-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of gypsogenin-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of gypsogenic acid-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises an increased content of a QS-21 adjuvant.

[00537] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a genetically modified plant disclosed herein

comprises a decreased content of Compound 7. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of Compound 8. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of Compound 9. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of Compound 10. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of Compound 11. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of glycyrrhizin (Compound 14). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of bayogenin (Compound 25). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of serjanic acid (Compound 26). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of serjanic acid-hexA-hex (Compound 32). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of soyasapogenol A (Compound 29). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of soyasapogenol B (Compound 30). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of soyasaponin VI (Compound 35). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of betavulgaroside IV (Compound 33). In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of hederagenin-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of gypsogenin-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of gypsogenic acid-3GlcA. In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of a QS-21 adjuvant.

[00538] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoid saponin and a decreased content of at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Comound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol

A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00539] In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Comopund 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Compound 7. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Compound 8. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Compound 9. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Compound 10. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Compound 11. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing glycyrrhizin (Compound 14). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing bayogenin (Compound 25). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing serjanic acid (Compound 26). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing serjanic acid-hexA-hex (Compound 32). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing soyasapogenol A (Compound 29). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing soyasapogenol B (Compound 30). In some embodiments, a method of

producing a triterpenoid saponin in a genetically modified cell comprises producing soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing soyasaponin VI (Compound 35). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing betavulgaroside IV (Compound 33). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing hederagenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing gypsogenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing gypsogenic acid-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified cell comprises producing a QS-21 adjuvant.

[00540] In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in a plant cell, a yeast cell, an algal cell, an insect cell, or a bacterium. In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in a plant cell. In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in a yeast cell. In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in an algal cell. In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in an insect cell. In some embodiments, the method of producing at least one triterpenoid saponin in a genetically modified cell comprises producing at least one triterpenoid saponins in a bacterium. In some embodiments, the plant cell is comprised in a plant or plant part.

[00541] In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant

comprises producing medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Compound 7. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Compound 8. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Compound 9. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Compound 10. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Compound 11. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing glycyrrhizin (Compound 14). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing bayogenin (Compound 25). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing serjanic acid (Compound 26). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing serjanic acid-hexA-hex (Compound 32). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing soyasapogenol A (Compound 29). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing soyasapogenol B (Compound 30). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing soyasaponin VI (Compound 35). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing betavulgaroside IV (Compound 33). In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing hederagenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing gypsogenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing gypsogenic acid-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in a genetically modified plant comprises producing a QS-21 adjuvant.

[00542] In some embodiments, a method of producing a triterpenoid saponin in an *in vitro* translation system comprises producing at least one triterpenoid saponin selected from any one of

medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Compound 7. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Compound 8. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Compound 9. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Compound 10. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Compound 11. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing glycyrrhizin (Compound 14). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing bayogenin (Compound 25). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing serjanic acid (Compound 26). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing serjanic acid-hexA-hex (Compound 32). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing soyasapogenol A (Compound 29). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing soyasapogenol B (Compound 30). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing soyasaponin VI (Compound 35). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing betavulgaroside IV

(Compound 33). In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing hederagenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing gypsogenin-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing gypsogenic acid-3GlcA. In some embodiments, a method of producing a triterpenoid saponin in an in vitro translation system comprises producing a QS-21 adjuvant.

[00543] In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of reducing the content of at least one triterpenoid saponin comprising medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Compound 7. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Compound 8. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Compound 9. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Compound 10. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Compound 11. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing glycyrrhizin (Compound 14). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing bayogenin (Compound 25). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing serjanic acid (Compound 26). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing serjanic acid-hexA-hex (Compound 32). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing soyasapogenol A (Compound 29). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing soyasapogenol B (Compound 30). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing

soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing soyasaponin VI (Compound 35). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing betavulgaroside IV (Compound 33). In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing hederagenin-3GlcA. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing gypsogenin-3GlcA. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing gypsogenic acid-3GlcA. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing a QS-21 adjuvant.

[00544] In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least one triterpenoid saponin selected from any one of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of increasing the content of at least one triterpenoid saponin comprising medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Compound 7. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Compound 8. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Compound 9. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Compound 10. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Compound 11. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing glycyrrhizin (Compound 14). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing bayogenin (Compound 25). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing serjanic acid (Compound 26). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing serjanic acid-hexA-hex (Compound 32). In some

embodiments, a method of increasing at least one triterpenoid saponin comprises increasing soyasapogenol A (Compound 29). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing soyasapogenol B (Compound 30). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing soyasaponin VI (Compound 35). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing betavulgaroside IV (Compound 33). In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing hederagenin-3GlcA. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing gypsogenin-3GlcA. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing gypsogenic acid-3GlcA. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing a QS-21 adjuvant.

[00545] In some embodiments, the content of both at least one triterpenoid saponin and at least one triterpenoid saponin biosynthetic intermediate are altered. In some embodiments, an at least one triterpenoid saponin and an at least one triterpenoid saponin biosynthetic intermediate are increased. In some embodiments, an at least one triterpenoid saponins and an at least one triterpenoid saponin biosynthetic intermediate are decreased. In some embodiments, an at least one triterpenoid saponin is increased and an at least one triterpenoid saponin biosynthetic intermediate is decreased. In some embodiments, an at least one triterpenoid saponin is decreased and an at least one triterpenoid saponin biosynthetic intermediate is increased. In some embodiments, the content of a triterpenoid saponin is altered without measurably altering the content of a triterpenoid saponin intermediate. In some embodiments, the content of a triterpenoid saponin intermediate is altered without measurably altering the content of a triterpenoid saponin.

[00546] The term "intermediate" may be used interchangeably in some embodiments with the term "biosynthetic intermediate", having all the same qualities and meanings.

[00547] In some embodiments, a biosynthetic intermediate of a triterpenoid saponin comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof.

[00548] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any

combination thereof. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 1. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 2. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising medicagenic acid (Compound 5). In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising oleanolic acid (Compound 3). In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising oleanolic acid-3GlcA. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising augustic acid (Compound 4). In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising augustic acid-3GlcA. In some embodiments, a genetically modified cell comprises an increased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising glycyrrhetinic acid (Compound 13).

[00549] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 1. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising Compound 2. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising medicagenic acid (Compound 5). In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising oleanolic acid (Compound 3). In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising oleanolic acid-3GlcA. In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising augustic acid (Compound 4). In some embodiments, a genetically modified plant comprises a decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising augustic acid-3GlcA. In some embodiments, a genetically modified plant comprises a

decreased content of at least one biosynthetic intermediate of a triterpenoid saponin comprising glycyrrhetinic acid (Compound 13).

[00550] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoid saponin biosynthetic intermediate and a decreased content of at least one triterpenoid saponin biosynthetic intermediate, said intermediate selected from any one of Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof.

[00551] In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least one triterpenoid saponin biosynthetic intermediate selected from any one of Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least Compound 1. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least Compound 2. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least medicagenic acid (Compound 5) In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least oleanolic acid (Compound 3) In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least oleanolic acid-3GlcA In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least augustic acid (Compound 4) In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least augustic acid-3GlcA In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing at least glycyrrhetinic acid (Compound 13).

[00552] In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least one triterpenoid saponin biosynthetic intermediate selected from any one of Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least Compound 1. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least Compound 2. In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least medicagenic acid (Compound 5) In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least oleanolic acid (Compound 3) In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least

oleanolic acid-3GlcA In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least augustic acid (Compound 4) In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least augustic acid-3GlcA In some embodiments, a method of increasing at least one triterpenoid saponin comprises increasing at least glycyrrhetinic acid (Compound 13).

[00553] In some embodiments, triterpenoid saponins biosynthetic intermediate comprises any Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or a combination thereof.

[00554] Unexpectedly, the present disclosure now shows that levels of triterpenoid saponins or their derivatives, metabolites, or biosynthetic intermediates can be increased in cells, for example plant cells or yeast cells or algal cells, or insect cells, or bacterium, or plants by genetically modifying the cell or plant to express at least one heterologous gene encoding an enzyme, for example an enzyme or enzymes of the triterpenoid saponin biosynthetic pathway. In other embodiments, described and exemplified herein are methods of genetically modifying an at least one endogenous gene in a plant cell, for example but not limited to a CSLG gene, to regulate expression, activity, or stability, or any combination thereof. The Examples below disclose enzyme activities and enzymes previously unknown to be part of the triterpenoid saponin biosynthetic pathway. Without this knowledge, production of the triterpenoid saponin compounds was not possible.

[00555] In some embodiments, the triterpenoid saponin metabolic pathway can result in cells or plants comprising elevated content of triterpenoid saponins, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or in plants having an increased content of these compounds in the plant or plant parts. In some embodiments, plants including but not limited to crop plants are produced, wherein the crop has an increased content of a useful triterpenoid saponin or triterpenoid saponins. In some embodiments, disclosed herein are the means and methods for producing cells including but not limited to plant cells, yeast, or algal cells; plants including but not limited to crop plants, or a part of a plant; having increased levels of a triterpenoid saponin, or triterpenoid saponins, or derivatives thereof, or metabolites thereof, or a biosynthetic intermediate thereof. Alternatively, or additionally, controlling the expression of genes disclosed herein may be used for the production of desired triterpenoid saponins for further use, for example in the pharmaceutical industry or for the formulation of dietary or other supplements, for example but not limited to sweeteners. In some embodiments, these high value saponins may be purified and used, e.g., as sweeteners, foaming agents, emulsifiers, preservatives, anti-carcinogens, hypocholesterolemic agents, anti-inflammatory agents, anti-oxidants, biological adjuvants, anti-microbial agents, insecticidal agents, anti-feedants,

or anti-fungal agents, or any combination thereof. The cells and plants disclosed herein comprise compounds of significant nutritional, pharmaceutical, and commercial value.

[00556] In some embodiments, a genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, as described herein, comprises an altered content of at least a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell or unmodified plant. In some embodiments, an altered content comprises an increased content. In some embodiments, for example, the genetically modified cell or genetically modified plant has an increased content of at least a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof.

[00557] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least one triterpenoid saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six triterpenoid saponins.

[00558] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a derivative of a triterpenoid saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two derivatives of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three derivatives of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four derivatives of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or

genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five derivatives of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six derivatives of a triterpenoid saponin or of triterpenoid saponins.

[00559] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a metabolite of a triterpenoid saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two metabolites of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three metabolites of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four metabolites of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five metabolites of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six metabolites of a triterpenoid saponin or of triterpenoid saponins.

[00560] In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least a biosynthetic intermediate of a triterpenoid saponin. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least two biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least three biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least four biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least five biosynthetic intermediates of

a triterpenoid saponin or of triterpenoid saponins. In some embodiments, the genetically modified cell or genetically modified plant or plant part comprising at least one genetically modified plant cell, has an increased content of at least six biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins.

[00561] As skilled artisan would recognize that the terms "content" and "level" in reference to a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, encompasses the quantity of the compound, for example the quantity of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof in a genetically modified cell or in a genetically modified plant, compared with a control cell or control plant. In this context, the terms "content" and "level" may be used interchangeably having all the same meanings and qualities.

[00562] In some embodiments, a triterpenoid saponin having increased content comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof. In some embodiments, a triterpenoid saponin having increased content comprises triterpenoid saponin selected from a medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a triterpenoid saponin having increased content comprises at least 1, 2, 3, 4, 5, 6, or more triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00563] In some embodiments, a triterpenoid saponin having increased content comprises at least 1, 2, 3, 4, 5, 6, or more triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-

3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00564] In some embodiments, a triterpenoid saponin having increased content in a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00565] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Compound 11. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Compound 7. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Compound 8. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Compound 9. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Compound 10. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in

a genetically modified plant or part thereof, comprises an increased content of glycyrrhizin (Compound 14). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of bayogenin (Compound 25). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of serjanic acid (Compound 26). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of serjanic acid-hexA-hex (Compound 32). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of soyasapogenol A (Compound 29). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of soyasapogenol B (Compound 30). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of soyasaponin VI (Compound 35). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of betavulgaroside IV (Compound 33). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of hederagenin-3GlcA. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of gypsogenin-3GlcA. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased

content of gypsogenic acid-3GlcA. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or in a genetically modified plant or part thereof, comprises an increased content of QS-21 adjuvant. A skilled artisan would appreciate that in some embodiments, instances wherein the content of one triterpenoid saponin is increased additional triterpenoid saponins, or intermediates, or a combination thereof may also be increased in the same time cell.

[00566] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of at least two biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins, said intermediates comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; , or any combination thereof.

[00567] In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bactierum, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising medicagenic acid (Compound 5). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 1. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 2. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising oleanolic acid (Compound 3). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising

oleanolic acid-3GlcA. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising augustic acid (Compound 4). In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising augustic acid-3GlcA. In some embodiments, a genetically modified cell, for example a plant cell or a yeast or an algal cell or an insect cell or a bacterium, or a genetically modified plant or part thereof, comprises an increased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising glycyrrhetinic acid (Compound 13), or any combination thereof.

[00568] In some embodiments, a triterpenoid saponin having decreased content in a plant or plant part comprising at least one genetically modified cell, comprises a compound having a bitter taste or a toxin. In some embodiments, a triterpenoid saponin having decreased content comprises triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a triterpenoid saponin having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00569] In some embodiments, a triterpenoid saponin having decreased content comprises at least 1, 2, 3, 4, 5, 6, or more triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11,

glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00570] In some embodiments, a triterpenoid saponin having decreased content in a genetically modified plant or part thereof comprising at least one genetically modified cell, comprises triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00571] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell comprises a decreased content of Compound 11. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of Compound 7. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of Compound 8. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of Compound 9. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of Compound 10. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of glycyrrhizin (Compound 14). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of bayogenin (Compound 25). In some embodiments, a genetically modified plant or part thereof comprising at

least one genetically modified plant cell, comprises a decreased content of bayogenin-hexA-hex-hex (Compound 31). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of serjanic acid (Compound 26). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of serjanic acid-hexA-hex (Compound 32). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of soyasapogenol A (Compound 29). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of soyasapogenol B (Compound 30). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of soyasaponin VI (Compound 35). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of betavulgaroside IV (Compound 33). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of hederagenin-3GlcA. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of gypsogenin-3GlcA. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of gypsogenic acid-3GlcA. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of QS-21 adjuvant. A skilled artisan would appreciate that in some embodiments, instances wherein the content of one triterpenoid saponin is decreased additional triterpenoid saponins, or intermediates, or a combination thereof may also be decreased in the same time cell.

[00572] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of at least two biosynthetic intermediates of a triterpenoid saponin or of triterpenoid saponins, said intermediates comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA,

augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof, or any combination thereof.

[00573] In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising medicagenic acid (Compound 5). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 1. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising Compound 2. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising oleanolic acid (Compound 3). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising oleanolic acid-3GlcA. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising augustic acid (Compound 4). In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising augustic acid-3GlcA. In some embodiments, a genetically modified plant or part thereof comprising at least one genetically modified plant cell, comprises a decreased content of a biosynthetic intermediate of a triterpenoid saponin, said intermediate comprising glycyrrhetinic acid (Compound 13), or any combination thereof.

[00574] In some embodiments, a genetically modified cell disclosed herein comprises an increased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00575] In some embodiments, a genetically modified cell disclosed herein comprises an increased

content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00576] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00577] In some embodiments, a genetically modified cell disclosed herein comprises a decreased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00578] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00579] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a decreased content of a phytosterol, or a derivative thereof, a

metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00580] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00581] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00582] In some embodiments, a method of producing at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises altering the content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00583] In some embodiments, a method of producing at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one triterpenoid saponin, a

derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00584] In some embodiments, a method of reducing the content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00585] In some embodiments, a method of reducing the content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00586] In some embodiments, a method of increasing the content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a

phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00587] In some embodiments, a method of increasing the content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

Plants, Plant Parts, and Cells

[00588] According to some embodiments, the cell or organism described herein, comprising the at least one heterologous gene encoding an enzyme has an elevated content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to a corresponding non-genetically modified cell or non-transgenic plant, respectively. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising an at least one modified endogenous gene encoding an enzyme of the steroidal alkaloid pathway, the steroidal saponin pathway, or the triterpenoid synthetic pathway has an elevated content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to a corresponding non-genetically modified plant. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising an at least one heterologous gene encoding an enzyme of the steroidal alkaloid synthetic pathway, the

steroidal saponin synthetic pathway, or the triterpenoid synthetic pathway has an elevated content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to a corresponding non-genetically modified plant. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising the at least one modified endogenous gene encoding an enzyme of the steroidal alkaloid synthetic pathway, the steroidal saponin synthetic pathway, or the triterpenoid synthetic pathway has a reduced content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to a corresponding non-genetically modified plant. As described throughout, the skilled artisan would appreciate that modifications of endogenous genes include but are not limited to increasing expression, decreasing expressing, or mutating the gene, or a combination thereof as described in detail herein.

[00589] According to some embodiments, the cell or organism described herein, comprising the at least one heterologous CSLG gene encoding a CSLG enzyme has an elevated content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, compared to a corresponding non-genetically modified cell or non-transgenic plant, respectively. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising an at least one modified endogenous CSLG gene encoding a CSLG enzyme of the steroidal alkaloid synthetic pathway, the steroidal saponin synthetic pathway, or the triterpenoid synthetic pathway has an elevated content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof compared to a corresponding non-genetically modified plant. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising an at least one heterologous CSLG gene encoding an enzyme of the steroidal alkaloid synthetic pathway, the steroidal saponin synthetic pathway, or the triterpenoid synthetic pathway has an elevated content of at least one steroidal alkaloid, derivative

thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof compared to a corresponding non-genetically modified plant. In other embodiments, the genetically modified plant or plant part comprising an at least one genetically modified cell described herein, comprising the at least one modified endogenous CSLG gene encoding an enzyme of the steroidal alkaloid synthetic pathway, the steroidal saponin synthetic pathway, or the triterpenoid synthetic pathway has a reduced content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof compared to a corresponding non-genetically modified plant.

[00590] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00591] In some embodiments, a cell or plant species or plant part that does not normally express for example at least a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG and/or does not normally produce a particular steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a particular steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a particular triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, is genetically modified to express a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof. In some embodiments, a cell or plant species or plant part that does not normally express a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG and/or does not normally produce a particular steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a particular steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a particular triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, is genetically modified to produce a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof. In other embodiments, a cell or plant species or plant part that normally expresses a given

amount of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof is genetically modified to overexpress a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof. In other embodiments, a cell or plant species or plant part that normally expresses a given amount of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof is genetically modified to reduce or silence expression of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof.

[00592] In other embodiments, a cell or plant species or plant part that normally produces a given amount of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, is genetically modified to overexpress a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof leading to an increases production of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof.

[00593] In other embodiments, a cell or plant species or plant part that normally produces a given amount of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, is genetically modified to reduce or silence expression of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG, or any combination thereof leading to a decreased production of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or any combination thereof.

[00594] In some embodiments, disclosed herein are seeds of the genetically modified plant, wherein plants grown from said seeds and expressing at least one saponin beta-amyrin synthase, cytochrome P450, a glycosyltransferase, acyltransferase, a glucuronosyltransferase, or CSLG

compared to plants grown from corresponding unmodified, thereby containing expression or overexpression of at least one saponin beta-amyrin synthase, cytochrome P450, a glycosyltransferase, acyltransferase, a glucuronosyltransferase, or CSLG and/or having increased content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00595] The cells disclosed herein comprise numerous varieties, including, but not limited to, yeast, algal, insect, bacterium, and plant cells. In some embodiments, a genetically modified cell comprises a plant cell. In some embodiments, a genetically modified cell comprises a yeast cell. In some embodiments, a genetically modified cell comprises an algal cell. In some embodiments, a genetically modified cell comprises an insect cell. In some embodiments, a genetically modified cell comprises a bacterium.

[00596] Generally, most yeast, algae, insect cells, and bacterium do not naturally synthesize steroidal alkaloids, steroidal saponins, or triterpenoid saponins. In some embodiments, in order to produce a steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a yeast or an algal cell or an insect cell or a bacterium, genetic modification of the yeast or algal cell or insect cell or bacterium comprises introduction of an enzyme or enzymes necessary to produce precursors or substrates, or a combination thereof, of the triterpenoid saponin biosynthetic pathway. For example, see **Example 14** below wherein the *S. cerevisiae* were genetically modified to express a heterologous UDP-glucose 6-dehydrogenase I in order to produce the necessary precursors for substrates of the triterpenoid saponin biosynthetic pathway. In some embodiments, a genetically modified cell expresses at least one heterologous gene for the production of precursors, substrates, or a combination thereof, of the steroidal alkaloid biosynthetic pathway, a steroidal saponin biosynthetic pathway, or the triterpenoid biosynthetic pathway. In some embodiments, a genetically modified cell expresses at least one heterologous gene for the production of precursors, substrates, or a combination thereof, of the steroidal alkaloid biosynthetic pathway, the steroidal saponin biosynthetic pathway, or the triterpenoid biosynthetic pathway, and expresses at least one heterologous gene encoding an enzyme of the steroidal alkaloid biosynthetic pathway, the steroidal saponin biosynthetic pathway, or the triterpenoid biosynthetic pathway.

[00597] In some embodiments, cells comprise a micro-organism. In some embodiments, the genetically modified cell is a yeast cell. In some embodiments, the yeast is from the *Saccharomycetes* order. In some embodiments, the yeast is from the *Saccharomycetes* order and is selected from the group of genera consisting of the *Saccharomyces* genus (e.g., *Saccharomyces cerevisiae*), the

Schizosaccharomyces genus, the *Pichia* genus (e.g., *Pichia pastoris*), the *Hansenula* genus (e.g., *Hansenula polymorpha*), the *Yarrowia* genus (e.g., *Yarrowia lipollytica*), the *Kluyveromyces* genus (e.g., *Kluyveromyces lactis*), and the *Candida* genus (e.g., *Candida albicans*, *Candida utilis*). In some embodiments, a yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus. In some embodiments, a yeast is selected from *Saccharomyces cerevisiae* or *Candida albicans.*

[00598] In some embodiments, a genetically modified yeast may be utilized in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin as described herein.

[00599] In some embodiments, the genetically modified cell is an algal cell. In some embodiments, the cell is any one of a variety of algae, including, but not limited to, chlorophytes (green algae), rhodophytes (red algae), or phaeo-phytes (brown algae). In some embodiments, the chlorophyte is from the *Chlamydomonadales* order or the *Chlorellales* order. In some embodiments, the chlorophyte is from the *Chlamydomonadales* order and is selected from the group of genera consisting of the *Chamydomonas* genus (e.g., *Chlamydomonas reinhardtii*) and the *Dunaliella* genus. In some embodiments, the chlorophyte is from the *Chlorellales* order and is selected from the *Chlorella* genus.

[00600] In some embodiments, a genetically modified alga may be utilized in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin as described herein.

[00601] In some embodiments, the genetically modified cell disclosed herein is a bacterial cell. In one embodiment, the bacterial cell is an *Escherichia coli* cell. In another embodiment, the cell is an *Acremonium rutilum*, *Aspergillus oryzae*, *Yarrowia lipolytica*, *Bacillus sp. JPJ*, *Brevundimonas sp. SGJ*, *E. herbicola*, *Citrobacter freundii*, *Symbiobacterium*, or *Pseudomonas aeruginosa* cell. In another embodiment, the bacterial cell is from a bacterium involved in fermentation of dairy products. In one embodiment, the bacterium is *Streptococcus lactis.* In another embodiment, the bacterium is a *Lactobacillus.* In one embodiment, the Lactobacillus is *Lactobacillus bulgaricus.* In another embodiment, the bacterium is a *Lactococcus* or a *Leuconostoc.*

[00602] In some embodiments, a genetically modified bacterium may be utilized in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin as described herein.

[00603] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a

derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00604] In some embodiments, a genetically modified plant disclosed herein comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a decreased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00605] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00606] In some embodiments, a genetically modified plant disclosed herein comprises a decreased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00607] In some embodiments, a method of producing at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00608] In some embodiments, a method of producing at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in a genetically modified cell or a genetically modified plant comprises increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00609] In some embodiments, a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate

thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00610] In some embodiments, a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising a reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, or a biosynthetic intermediate thereof, and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00611] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00612] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, comprises genetically modifying at least one plant cell or at least one cell of a plant or plant part, said genetically modified plant cell comprising an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and a reduced content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00613] In some embodiments, the genetically modified cell is a plant cell. Plant cells can be obtained from a wide range of plants, as discussed below. The plant cell may be a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a bract cell, a callus cell, and a flower cell. In some embodiments, a genetically modified cell may then be comprised in a plant leaf, in a plant petiole, in a plant stem or stalk, in a plant root, in a plant bud, in a plant tuber, in a plant bean, in a plant grain or kernel, in a plant fruit, in a plant nut, in a plant legume, in a plant seed, in a plant bract, in a plant callus, or in a plant flower.

[00614] In some embodiments, the genetically modified plant cell is grown as such in culture, independent of a plant or plant part. Methods for culturing plants and plant parts are well known in the art, for example see Ochoa-Villarreal M, Howat S, Hong S, *et al*. Plant cell culture strategies for the production of natural products. BMB Rep. 2016;49(3):149–158, which is incorporated herein in

its entirety.

[00615] In some embodiments, the genetically modified plant cell is grown as such in culture, independent of a plant or plant part and is utilized in a method provided herein to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins. Suspensions of genetically modified cells and tissue cultures derived from the genetically modified cells are also encompassed within the scope described herein. The cell suspension and tissue cultures can be used for the production of a desired steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a desired steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a desired triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and, which are then extracted from the cells or the growth medium. Alternatively, the genetically modified plant cell and/or tissue culture are used for regenerating a transgenic plant having modified expression or overexpression of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG , therefore expressing or overexpressing a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase 1, and a cellulose synthase like G (CSLG) in a cell or in a plant, as compared to a corresponding unmodified cell or plant, respectively. Therefore, having modified content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00616] In some embodiments, the genetically modified plant cell is comprised in a plant. In some embodiments, the genetically modified plant cell is comprised in a plant, wherein the plant is utilized in a method provided herein to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins. In some embodiments, the genetically modified plant cell is comprised in a plant tissue. In some embodiments, the genetically modified plant cell is comprised in a plant tissue, wherein the plant tissue is utilized in a method provided herein to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins. In some embodiments, the genetically modified plant cell is comprised in a plant organ. In some embodiments, the genetically modified plant cell is comprised in a plant organ, wherein the plant organ is utilized in a method provided herein to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins. In some embodiments, the genetically modified plant cell is comprised in a plant part. In some embodiments, the genetically modified plant cell is comprised in a plant part, wherein the plant part is utilized in a method provided herein to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00617] In some embodiments, a genetically modified cell comprises a plant cell comprising a leaf

cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a bract cell, a callus cell, or a flower cell. In some embodiments, a genetically modified cell comprises a plant cell comprised in a plant leaf, in a plant petiole, in a plant stem or stalk, in a plant root, in a plant bud, in a plant tuber, in a plant bean, in a plant grain or kernel, in a plant fruit, in a plant nut, in a plant legume, in a plant seed, in a plant bract, in a plant callus, or in a plant flower, or in a combination thereof. In some embodiments, a genetically modified plant cell is comprised within a plant part, wherein said plant part comprises a leaf , a petiole, a plant stem or stalk, a root, a bud, a tuber, a bean, a grain or kernel, a fruit, a nut, a legume, a seed or seed, a bract, a callus, or a flower.

[00618] In some embodiments, a genetically modified plant cell may be utilized in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin as described herein.

[00619] In some embodiments, a genetically modified cell comprises a genetically modified plant cell (e.g., cultured plant cells). In other embodiments, a genetically modified plant cell is comprised in plant or a plant organ(s), plant tissue(s), or plant part(s). In some embodiments, the term "plant" may be used interchangeably with the term "plant part" having all the same meanings and qualities.

[00620] In some embodiments, a genetically modified plant cell, plant, plant organ, plant tissue, or plant part may be utilized in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin as described herein.

[00621] The content of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof present in a genetically modified cell or plant or plant part is measured as exemplified hereinbelow and as is known to a person skilled in the art.

[00622] In some embodiments, an offspring plant comprises increased contents of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to at least one of the progenitor plants.

[00623] In some embodiments, a plant or plant cell or plant tissue or plant part as disclosed herein is selected from the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Poales* order, the *Lamiales* order, the *Cucurbitales* order, the *Asterales* order, the *Malpighiales* order, the *Rosales* order, the *Fagales* order, the *Sapindales* order, the *Arecales* order, the *Ericales* order, the *Gentianales* order, the *Ranunculales* order, the *Zingiberales* order, the *Saxifragales* order, the *Vitales* order, the *Pinales* order, the *Cornales* order, the *Asparagales* order, the *Dioscoreales* order, and the *Liliales* order. In some embodiments, a plant

or plant cell or plant tissue or plant part as disclosed herein comprises a cell from a plant selected from the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Poales* order, the *Lamiales* order, the *Cucurbitales* order, the *Asterales* order, the *Malpighiales* order, the *Rosales* order, the *Fagales* order, the *Sapindales* order, the *Arecales* order, the *Ericales* order, the *Gentianales* order, the *Ranunculales* order, the *Zingiberales* order, the *Saxifragales* order, the *Vitales* order, the *Pinales* order, the *Cornales* order, the *Asparagales* order, the *Dioscoreales* order, and the *Liliales* order.

[00624] In some embodiments, a plant or plant cell or plant tissue or plant part as disclosed herein comprises a cell from a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

[00625] In some embodiments, a plant or plant cell or plant tissue or plant part as disclosed herein is in the *Caryophyllales* order and is selected from the group of genera consisting of the *Spinacia* genus (e.g., *Spinacia oleracea*), the *Chenopodium* genus (e.g., *Chenopodium quinoa*), the *Beta* genus (e.g., *Beta vulgaris*) the *Rheum* genus (e.g., *Rheum hybridum, Rheum rhaponticum, Rheum rhabarbarum, Rheum ribes*), the *Vaccaria* genus (e.g., *Vaccaria hispanica*), the Saponaria genus (e.g., *Saponaria vaccaria*), and the *Gypsophila* genus (e.g., *Gypsophila paniculata*). In some embodiments, a *Caryophyllales* plant is selected from the group consisting of spinach, beetroot, and quinoa.

[00626] In some embodiments, a plant or plant cell or plant tissue or plant part as disclosed herein is in the *Solanales* order and is selected from the group of genera consisting of the *Nicotiana* genus (e.g., *Nicotiana benthamiana*), the *Solanum* genus (e.g., *Solanum lycopersicum, Solanum tuberosum, Solanum melongena, Solanum pennellii, Solanum chacoense, Solanum dulcamara*), the *Capsicum* genus (e.g., *Capsicum annuum*), the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus. In some embodiments, a plant as disclosed herein comprises a *Solanaceae* crop plant. In some embodiments, a *Solanaceae* crop plant is selected from the group consisting of *Solanum lycopersicum, Solanum pennellii, Solanum tuberosum, Solanum chacoense, Capsicum annuum, Solanum dulcamara*, and *Solanum melongena*. In some embodiments, a *Solanaceae* or *Solanales* plant is selected from the group consisting of ground cherry, eggplant, potato, tomato, wild tomato, potato, wild potato, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, tobacco, and bittersweet.

[00627] In some embodiments, the plant is in the *Fabales* order and is selected from the group of genera consisting of the *Glycyrrhiza* genus (e.g., *Glycyrrhiza uralensis, Glycyrrhiza glabra* [licorices]), the *Medicago* genus (e.g., *Medicago sativa, Medicago truncatula*), the *Quillaja* genus

(e.g., *Quillaja saponaria*), the *Glycine* genus (e.g., *Glycine max* [soy/soybean]), the *Lotus* genus (e.g., *Lotus japonicus*), the *Cicer* genus (e.g., *Cicer arietinum* [chickpea, garbanzo bean]), the *Phaseolus* genus (e.g., *Phaseolus vulgaris* [string bean, common bean, French bean]), the *Pisum* genus (e.g., *Pisum sativum* [pea]), the *Arachis* genus (e.g., *Arachis hypogaea* [peanut]), the *Lupinus* genus (e.g., *Lupinus albus* [lupin/lupine]), and the *Acacia* genus. In some embodiments, a *Fabales* plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

[00628] In some embodiments, the plant is from the *Malvales* order and is selected from the *Theobroma* genus (e.g., *Theobroma cacao*).

[00629] In some embodiments, the plant is from the *Apiales* order and is selected from the group of genera consisting of the *Daucus* genus (e.g., *Daucus carota*), the *Apium* genus (e.g., *Apium graveolens),* the *Petroselinum* genus (e.g., *Petroselinum crispum*), the *Panax* genus (e.g., *Panax ginseng*), the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus (e.g., *Centella asiatica).*

[00630] In some embodiments, the plant is from the *Brassicales* order and is selected from the group of genera consisting of the *Arabidopsis* genus (e.g., *Arabidopsis thaliana*), the *Brassica* genus (e.g., *Brassica oleracea* [cabbages], *Brassica juncea* [white mustard], *Brassica nigra* [black mustard], *Brassica napus*), the *Capparis* genus (e.g., *Capparis spinosa* [caper]), and the *Carica* genus (e.g., *Carica papaya* [papaya]).

[00631] In some embodiments, the plant is from the *Poales* order and is selected from the group of genera consisting of the *Oryza* genus (e.g., *Oryza sativa* and *Oryza glaberrima* [rice]), the *Hordeum* genus (e.g., *Hordeum vulgare* [barley]), the *Avena* genus (e.g., *Avena sativa* [oat], *Avena strigosa*), and the *Triticum* genus (e.g., *Triticum spelta* [spelt]).

[00632] In some embodiments, the plant is from the *Lamiales* order and is selected from the group of genera consisting of the *Salvia* genus (e.g., *Salvia hispanica* [chia]), the *Sesamum* genus (e.g., *Sesamum indicum* [sesame, benne]), and the *Olea* genus (e.g., *Olea europaea* [olive]).

[00633] In some embodiments, the plant is from the *Cucurbitales* order and is selected from the *Cucurbita* genus (e.g., squash/pumpkin, including, but not limited to, *Cucurbita pepo, Cucurbita maxima, Cucurbita argyrosperma, or Cucurbita moschata*).

[00634] In some embodiments, the plant is from the *Asterales* order and is selected from the group of genera consisting of the *Helianthus* genus (e.g., *Helianthus annuus* [sunflower], *Helianthus verticallatus* [whorled sunflower], *Helianthus tuberosus* [Jerusalem artichoke]), the *Artemesia* genus (e.g., *Artemesia annua*), the *Galatella* (Aster) genus (e.g., *Galatella sedifolia*), and the *Taraxacum* genus (e.g., *Taraxacum officinale* [dandelion]).

[00635] In some embodiments, the plant is from the *Malpighiales* order and is selected from the group of genera consisting of the *Linum* genus (e.g., *Linum usitatissimum* [flax, linseed]), the

Bruguiera genus (e.g., *Bruguiera gymnorhiza*), the *Euphorbia* genus (e.g., *Euphorbia tirucalli*), the *Ricinus* genus (e.g., *Ricinus communis* [castor]), the *Kandelia* genus (e.g., *Kandelia candel),* and the *Rhizophora* genus (e.g., *Rhizophora stylosa*).

[00636] In some embodiments, the plant is from the *Rosales* order and is selected from the group of genera consisting of the *Prunus* genus (e.g., *Prunus dulcis* [almond], *Prunus amygdalus*) and the *Cannabis* genus (e.g., hemp, including *Cannabis sativa*).

[00637] In some embodiments, the plant is from the *Fagales* order and is selected from the group of genera consisting of the *Corylus* genus (e.g., hazel/hazelnut/cobnut/filbert nut, including, but not limited to, *Corylus avellana*), the *Betula* genus (e.g., Betula pendula [silver birch], *Betula pubescens* [white or downy birch], *Betula platyphylla*) and the *Juglans* genus (e.g., *Juglans regia* [Persian or English walnut], *Juglans nigra* [black walnut], *Juglans cinera* [butternut]).

[00638] In some embodiments, the plant is from the *Sapindales* order and is selected from the group of genera consisting of the *Anacardium* genus (e.g., *Anacardium occidentale* [cashew]), the *Pistacia genus* (e.g., *Pistacia vera* [pistachio]), the *Citrus* genus (numerous species and hybrids), the *Aesculus* genus (e.g., *Aesculus hippocastanum, Aesculus turbinata*), and the *Peganum* genus.

[00639] In some embodiments, the plant is from the *Arecales* order and is selected from the *Cocus* genus (e.g., *Cocus nucifera* [coconut]).

[00640] In some embodiments, the plant is from the *Ericales* order and is selected from the *Maesa* genus.

[00641] In some embodiments, the plant is from the *Gentianales* order and is selected from the group of genera consisting the *Nerium* genus (e.g., *Nerium oleander*), the *Gentianamacrophylla* genus (e.g., *Gentianamacrophylla straminea*), the *Catharanthus* genus (e.g., *Catharanthus roseus*), the *Rauwolfia* genus, and the *Cinchona* genus.

[00642] In some embodiments, the plant is from the *Ranunculales* order and is selected from the genera consisting of the *Nigella* genus (e.g., *Nigella sativa*), the *Papaver* genus, the *Eschscholtzia* genus (e.g., *Eschscholtzia californica*), the *Coptis* genus, the *Berberis* genus, and the *Thalictrum* genus.

[00643] In some embodiments, the plant is from the *Zingiberales* order and is selected from the *Cheilocostus* genus (e.g., *Cheilocostus speciosus*).

[00644] In some embodiments, the plant is from the *Saxifragales* order and is selected from the *Kalanchoe* genus (e.g., *Kalanchoe daigremontiana*).

[00645] In some embodiments, the plant is from the *Vitales* order and is selected from the *Vitis* genus (e.g., *Vitis vinifera* [grape]).

[00646] In some embodiments, the plant is from the *Pinales* order and is selected from the *Taxus*

genus (e.g., *Taxus brevifolia, Taxus baccdata, Taxus cuspidata, Taxus canadensis, Taxus floridana*).

[00647] In some embodiments, the plant is from the *Cornales* order and is selected from the *Camptotheca* genus (e.g., *Camptotheca acuminata*). In some embodiments, the plant is from the *Asparagales* order and is selected from the *Agave* genus (e.g., *Agave americana*, *Agave attenuata*, *Agave tequilana*), the *Asparagus* genus (e.g., *Asparagus officinalis*), or the *Yucca* genus (e.g., *Yucca filamentosa*). In some embodiments, the plant is from the *Dioscoreales* order and is selected from the *Borderea* genus, the *Dioscorea* genus, the *Epipetrum* genus, the *Rajania* genus, the *Stenomeris* genus, or the *Tamus* genus. In some embodiments, the plant is from the *Liliales* order and is selected from the *Liliaceae* family (e.g., *Clintonia borealis*, *Nomocharis aperta*, *Calochortus catalinae*, *Streptopus lariceolatus*).

[00648] A skilled artisan would appreciate that plant breeding can be accomplished through many different techniques ranging from simply selecting plants with desirable characteristics for propagation, to methods that make use of knowledge of genetics and chromosomes, to more complex molecular techniques.

[00649] A skilled artisan would appreciate that the term "hybrid plant" may encompass a plant generated by crossing two plants of interest, propagating by seed or tissue and then growing the plants. When plants are crossed sexually, the step of pollination may include cross pollination or self-pollination or back crossing with an untransformed plant or another transformed plant. Hybrid plants include first generation and later generation plants.

Biological Activity

[00650] In some embodiments, the genetically modified cell or genetically modified plant comprises an endogenous enzyme comprising the same or similar activity with an enzyme encoded by the at least one heterologous gene. In other embodiments, the genetically modified cell or genetically modified plant does not comprise an endogenous enzyme comprising the same or similar activity with an enzyme encoded by the at least one heterologous gene. In some embodiments, the genetically modified plant cell or plant comprises an endogenous enzyme comprising altered activity of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin biosynthesis enzyme.

[00651] In some embodiments, the biological activity of at least one saponin beta-amyrin synthase, cytochrome P450, a glycosyltransferase, acyltransferase, a glucuronosyltransferase, or CSLG, or combination thereof encoded by the heterologous gene is altered compared with an endogenous saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG protein present in the genetically modified cell or genetically modified plant. In some embodiments, the biological activity of at least one endogenous saponin

beta-amyrin synthase, cytochrome P450, a glycosyltransferase, acyltransferase, a glucuronosyltransferase, or CSLG, or combination thereof is altered compared with the endogenous saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG protein present in the non-genetically modified cell or non-genetically modified plant.

[00652] A skilled artisan would recognize that the term "biological activity" refers to any activity associated with a protein that can be measured by an assay. In some embodiments, the biological activity of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG comprises an enzyme activity necessary for the biosynthesis of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof in a cell or a plant. In some embodiments, the biological activity of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a glucuronosyltransferase, or a CSLG affects the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least a part of a plant. In some embodiments, an altered biological activity comprises increased enzyme activity. In some embodiments, an altered biological activity comprises decreased enzyme activity. In some embodiments, an altered biological activity comprises increased stability of the polypeptide. In some embodiments, an altered biological activity comprises decreased stability of the polypeptide. In some embodiments, an altered biological activity comprises increased expression so higher levels of the enzyme polypeptide. In some embodiments, an altered biological activity comprises decreased expression of the enzyme polypeptide.

[00653] In some embodiments, the biological activity of an enzyme is altered compared with a control enzyme.

[00654] In some embodiments, the altered biological activity comprises (a) increased enzyme activity of the saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or increased stability of the saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or decreased enzyme activity of the saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or decreased stability of the saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase,

glucuronosyltransferase, or CSLG.

[00655] In some embodiments, the biological activity of a saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG is increased. In some embodiments, the biological activity of a saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG is decreased. In some embodiments, a saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG has increased stability. In some embodiments, a saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG has decreased stability.

[00656] In some embodiments, the altered biological activity comprises increased enzyme activity of said at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or increased stability of said at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or decreased enzyme activity of said at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; or decreased stability of said at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; compared to the biological activity in the endogenous enzyme, if present, in said genetically modified cell or genetically modified plant.

Transgenic Plants, Transgenic Yeast, Transgenic Algae, Transgenic Insect Cells, Transgenic Bacterium

[00657] Cloning of a polynucleotide encoding an enzyme encoded by the at least one heterologous gene, wherein the enzyme comprises at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG can be performed by any method as is known to a person skilled in the art. Cloning of a polynucleotide encoding an at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG as described herein, can be performed by any method as is known to a person skilled in the art. In certain embodiments, a polynucleotide encoding a CSLG enzyme is cloned. Various DNA constructs may be used to express the desired gene in a desired cell or organism. In some embodiments, various DNA constructs are used to expression a CSLG gene in a cell. In some embodiments, various DNA constructs are used to expression a CSLG gene in a plant cell, a yeast cell, an insect cell, an algal cell, or a bacterium. In some embodiments, various DNA constructs are used to expression a CSLG gene in a plant cell. In some embodiments, various DNA

constructs are used to expression a CSLG gene in a yeast cell. In some embodiments, various DNA constructs are used to expression a CSLG gene in a plant cell or a yeast cell. In some embodiments, various DNA constructs are used to expression a CSLG gene in an algal cell. In some embodiments, various DNA constructs are used to expression a CSLG gene in an insect cell. In some embodiments, various DNA constructs are used to expression a CSLG gene in a bacterium. In some embodiments, various DNA constructs are used to expression a heterologous CSLG gene in the cell. In some embodiments, when the cell is a plant cell the CSLG gene is a heterologous gene. In some embodiments, when the cell is a plant cell the CSLG gene is a homologous gene, wherein said expression is altered, for example increased or decreased, or said CSLG gene is mutated to alter the activity of the CSLG enzyme. In some embodiments, when the cell is a plant cell, the plant cell is comprised within a plant or a plant part.

[00658] According to certain embodiments, the gene may comprise part of an expression vector comprising all necessary elements for expression of the gene and optional regulatory components.

[00659] According to certain embodiments, the expression is controlled by a constitutive promoter. According to other embodiments, the expression is controlled by a transient promoter. According to certain embodiments, the constitutive promoter is specific to a cell or to a plant or plant tissue. According to some embodiments, the tissue specific promoter is selected from the group consisting of root, tuber, leaves and fruit specific promoter. Root specific promoters are described, e.g. in Martinez, E. et al. 2003. Curr. Biol. 13:1435-1441. Fruit specific promoters are described among others in Estornell L.H et al. 2009. Plant Biotechnol. J. 7:298-309 and Fernandez A. I. Et al. 2009 Plant Physiol. 151:1729-1740. Tuber specific promoters are described, e.g. in Rocha-Sosa M, et al., 1989. EMBO J. 8:23-29; McKibbin R.S. et al., 2006. Plant Biotechnol J. 4(4):409-18. Leaf specific promoters are described, e.g. in Yutao Yang, Guodong Yang, Shijuan Liu, Xingqi Guo and Chengchao Zheng. Science in China Series C: Life Sciences. 46: 651-660. Accordingly, in some embodiments, a promoter comprises a yeast specific promoter. Accordingly, in some embodiments, a promoter comprises an algal specific promoter.

[00660] According to certain embodiments, the expression vector further comprises regulatory elements at the 3' non-coding sequence. As used herein, the "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. For plants, the use of different 3' non-coding sequences is exemplified by Ingelbrecht I L et al. (1989. Plant Cell 1:671-680).

[00661] Those skilled in the art will appreciate that the various components of the nucleic acid

sequences and the transformation vectors described herein are operatively linked, so as to result in expression of said nucleic acid or nucleic acid fragment. Techniques for operatively linking the components of the constructs and vectors described herein are well known to those skilled in the art. Such techniques include the use of linkers, such as synthetic linkers, for example, including one or more restriction enzyme sites.

[00662] One skilled in the art would appreciate that the term "operably linked" may encompass the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is regulated by the other. For example, a promoter is operably linked with a coding sequence when it is capable of regulating the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation.

[00663] Methods for transforming a plant according to the teachings disclosed herein are known to those skilled in the art. As used herein the term “transformation” or “transforming” describes a process by which a foreign DNA, such as a DNA construct, including expression vector, enters and changes a recipient cell into a transformed, genetically altered or transgenic cell. Transformation may be stable, wherein the nucleic acid sequence is integrated into the organism genome and as such represents a stable and inherited trait, or transient, wherein the nucleic acid sequence is expressed by the cell transformed but is not integrated into the genome, and as such represents a transient trait. According to some embodiments the nucleic acid sequence disclosed herein is stably transformed into the cell.

[00664] The genetically altered cells or plants having altered content of the desired at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG; and altered content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof. In some embodiments, genetically modified cells or plants produced as described herein are typically first selected based on the expression of the gene or protein. Cells or plants having expressing the enzyme encoding from the at least one heterologous gene may be identified and , are then analyzed for the content of at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG and/or at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof .

[00665] In some embodiments, the genetically altered plants having altered content of the desired at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, according to the teachings herein are first selected based on the expression of the gene or protein. Plants having enhanced or aberrant expression of the gene or protein, are then analyzed for the content of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof. In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00666] Detection of at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG gene and/or at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is performed employing standard methods of molecular genetics, known to a person of ordinary skill in the art. Similarly, purification and/or detection of at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof is performed employing standard methods of molecular genetics or protein chemistry known to a person of ordinary skill in the art.

[00667] For measuring the gene(s), cDNA or mRNA should be obtained from a cell or a plant in which the nucleic acid is expressed. The sample may be further processed before the detecting step. For example, the polynucleotides in the cell or tissue sample may be separated from other components of the sample, may be amplified, etc. All samples obtained from an organism, including those subjected to any sort of further processing are considered to be obtained from the organism.

[00668] For measuring the gene(s) or silencing molecule(s) expression, cDNA or mRNA should be obtained from an organ in which the nucleic acid is expressed. The sample may be further processed before the detecting step. For example, the polynucleotides in the cell or tissue sample may be separated from other components of the sample, may be amplified, etc. All samples obtained from an organism, including those subjected to any sort of further processing are considered to be obtained

from the organism.

[00669] Detection of the gene(s) in some embodiments, requires amplification of the polynucleotides taken from the candidate genetically modified cell or genetically modified plant or part thereof. In some embodiments a plant part comprises a plant organ or a plant tissue.

[00670] Methods for DNA amplification are known to a person skilled in the art. Most commonly used method for DNA amplification is PCR (polymerase chain reaction; see, for example, PCR Basics: from background to Bench, Springer Verlag, 2000; Eckert et al., 1991. PCR Methods and Applications 1:17). Additional suitable amplification methods include the ligase chain reaction (LCR), transcription amplification and self-sustained sequence replication, and nucleic acid-based sequence amplification (NASBA).

[00671] According to certain embodiments, the nucleic acid sequence comprising the at least one saponin beta-amyrin synthase, cytochrome P450, glycosyltransferase, acyltransferase, glucuronosyltransferase, or CSLG, or any combination thereof may in some embodiments, further comprises a nucleic acid sequence encoding a selectable marker. According to certain embodiments, the selectable marker confers resistance to antibiotic or to an herbicide; in these embodiments the transgenic cells or plants are selected according to their resistance to the antibiotic or herbicide.

Breeding

[00672] In some embodiments, transformation techniques including breeding through transgene editing, use of transgenes, use of transient expression of a gene or genes, or use of molecular markers, or any combination thereof, may be used in the breeding of a plant having an altered expression. If transformation techniques require use of tissue culture, transformed cells may be regenerated into plants in accordance with techniques well known to those of skill in the art. The regenerated plants may then be grown and crossed with the same or different plant varieties using traditional breeding techniques to produce seed, which are then selected under the appropriate conditions.

[00673] The content of steroidal alkaloids and/or steroidal saponins is measured as exemplified hereinbelow and as is known to a person skilled in the art.

[00674] In some embodiments, an offspring plant comprises decreased anti-nutritional contents or decreased toxins compared to at least one of the progenitor plants. In some embodiments, an offspring plant comprises improved resistance to a plant pathogen, pest, or predator compared to at least one of the progenitor plants.

[00675] In some embodiments, a plant as disclosed herein comprises a crop plant from any of the orders listed in the previous section. In some embodiments, a plant as disclosed herein comprises a crop plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the

Malvales order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

[00676] In some embodiments, when the plant comprises a crop plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus.

[00677] In some embodiments, when a plant is a crop plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus. In some embodiments, the *Caryophyllales* plant is selected from the group consisting of spinach, beetroot, and quinoa.

[00678] In some embodiments, when a plant is a crop plant in the *Solanales* order, said plant from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus. In some embodiments, when a plant is a crop plant in the *Solanales* order, said plant is selected from the group consisting of *Solanum lycopersicum*, *Solanum pennellii*, *Solanum tuberosum*, *Solanum chacoense*, *Capsicum annuum*, *Solanum melongena*, *Solanum dulcamara,* and *Nicotiana benthamiana*. In some embodiments, the *Solanales* plant is selected from the group consisting of ground cherry, eggplant, potato, tomato, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, tobacco, and bittersweet. In some embodiments, the *Solanales* plant comprises a *Solanaceae* plant.

[00679] In some embodiments, when a plant is a crop plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus. In some embodiments, the *Fabales* plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

[00680] In some embodiments, when a plant is a crop plant in the *Malvales* order, said plant is selected from the *Theobroma* genus.

[00681] In some embodiments, when a plant is a crop plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium* genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus.

[00682] In some embodiments, when a plant is a crop plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

[00683] In some embodiments, the plant is a crop plant of any order, family, genus, or species disclosed herein.

[00684] A skilled artisan would appreciate that plant breeding can be accomplished through many different techniques ranging from simply selecting plants with desirable characteristics for propagation, to methods that make use of knowledge of genetics and chromosomes, to more complex molecular techniques.

[00685] A skilled artisan would appreciate that the term "hybrid plant" may encompass a plant generated by crossing two plants of interest, propagating by seed or tissue and then growing the plants. When plants are crossed sexually, the step of pollination may include cross pollination or self-pollination or back crossing with an untransformed plant or another transformed plant. Hybrid plants include first generation and later generation plants. Disclosed herein is a method to manipulate and improve a plant trait, for a non-limiting example - increasing plant resistance, decreasing anti-nutritional properties in a plant, decreasing toxins or bitter tasting compounds in a plant, increasing pharmaceutical properties in a plant, or any combination thereof.

Biomarkers

[00686] A skilled artisan would appreciate that the term "biomarker" comprises any measurable substance in an organism whose presence is indicative of a biological state or a condition of interest. In some embodiments, the presence of a biomarker is indicative of the presence of a compound or a group of compounds of interest. In some embodiments, the concentration of a biomarker is indicative of the concentration of a compound or a group of compounds of interest. In some embodiments, the concentration of a biomarker is indicative of an organism phenotype.

[00687] CSLG enzymes are hereby disclosed to have an essential role in the biosynthesis of steroidal alkaloids, steroidal saponins, and triterpenoid saponins found in many plants. Thus, in some embodiments, the expression level of CSLG is indicative of the capacity of a plant to produce steroidal alkaloids, or derivatives, metabolites, or biosynthetic intermediates thereof; steroidal saponins, or derivatives, metabolites, or biosynthetic intermediates thereof; triterpenoid saponins, or derivatives, metabolites, or biosynthetic intermediates thereof; or combinations thereof.

Methods for Producing Steroidal Alkaloids, Steroidal Saponins, and Triterpenoid Saponins

[00688] Provided herein are methods of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin in a genetically modified cell or plant comprising a genetically modified plant cell or a plant part comprising a genetically modified plant part, the methods comprising: introducing a polynucleotide sequence into said cell, wherein said polynucleotide sequence is optionally comprised in a vector, wherein said polynucleotide sequence comprises at least one heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin

synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, and a cellulose synthase like G (CSLG); and expressing said at least one heterologous gene in said cell. In some embodiments, the polynucleotide sequence introduced comprises multiple polynucleotide sequences, wherein each polynucleotide sequence comprises at least one heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, and a cellulose synthase like G (CSLG). In some embodiments, the polynucleotide sequence introduced comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, or more polynucleotide sequences each polynucleotide sequence comprising at least one heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, and a cellulose synthase like G (CSLG). In some embodiments, the polynucleotide sequence introduced comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, or more polynucleotide sequences each comprises only one heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, and a cellulose synthase like G (CSLG). In some embodiments, the polynucleotide sequence introduced comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, or more polynucleotide sequences each comprising any combination of 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, or more heterologous genes encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, and a cellulose synthase like G (CSLG). A skilled artisan would appreciate that the polynucleotide sequence or sequences are introduced into a cell where they are expressed, resulting in expression of the encoded enzyme or enzymes in the genetically modified cell.

[00689] In some embodiments, disclosed herein is a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin in a genetically modified cell, the method comprising:

(a) introducing an at least one heterologous gene into said cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said heterologous gene is optionally comprised in a vector; and

(b) expressing said at least one heterologous gene in said cell;

wherein said cell comprises an increased content of at least one steroidal alkaloid, at least one steroidal saponin, or at least one triterpenoid saponin compared to a corresponding unmodified cell.

[00690] In some embodiments, said introducing further comprising introducing an at least one additional heterologous gene into said cell, said heterologous gene selected from the group consisting of the group encoding a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an

acyltransferase, and a UDP-glucose 6-dehydrogenase I, or any combination thereof, wherein said at least one additional heterologous gene is optionally comprised in a vector; and further comprising expressing said at least one additional heterologous gene in said cell. In some embodiments, said at least one heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or said at least one additional heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or a combination thereof of (a) and (b).

[00691] In some embodiments, said introducing comprises transforming said at least one cell with said at least one heterologous gene or a polynucleotide sequence encoding said at least one heterologous gene, or the vector comprising said at least one heterologous gene; or said at least one additional heterologous gene or a polynucleotide sequences encoding said at least one additional heterologous gene, or the vector comprising said at least one additional heterologous gene; or a combination thereof of (a) and (b); wherein said expressing comprises transient expression or constitutive expression.

[00692] Cellulose synthase like G nucleic acid sequences encoding a CSLG enzyme are described in detail throughout and exemplified in the Examples. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. Cellulose synthase like G nucleic acid sequences encoding a CSLG enzyme are described in detail throughout and exemplified in the Examples. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00693] Cellulose synthase like G nucleic acid sequences encoding a CSLG enzyme are described in detail throughout and exemplified in the Examples. In some embodiments, the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00694] In some embodiments, in a method of producing a triterpenoid saponin, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39,

41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104. In some embodiments, in a method of producing a triterpenoid saponin, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104.

[00695] In some embodiments, in a method of producing a triterpenoid saponin, the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104.

[00696] The GAME proteins, GAME1 and GAME4, were used as baits in potato or tomato, as described herein in the Examples. Genes that were co-expressed with GAME1 or GAME4 are shown in **Tables 5-8**, and the enzymes of the steroidal alkaloid pathway, along with a portion of the enzymes of steroidal saponin pathway, are presented in **FIGURE 1** and described in detail throughout and exemplified in the Examples. The SOAP nucleic acid sequences respectively encoding the SOAP enzymes of the triterpenoid saponin pathway presented in **FIGURE 20A**, are described in detail throughout and exemplified in the Examples.

[00697] In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the nucleic acid sequence encoding said at least one additional heterologous gene encodes a β-amyrin synthase, said nucleic acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or an acyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or any combination thereof.

[00698] In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the amino acid sequence of said encoded at least one additional heterologous gene encodes a β-amyrin synthase, said amino acid sequence set forth in SEQ ID NO: 48; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 48; or a cytochrome P450, said amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or a glycosyl transferase, said amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or an acyltransferase, said amino acid sequence set forth in SEQ ID NO: 64; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 64; or a UDP-glucose 6-dehydrogenase 1, said amino acid sequence set forth in SEQ ID NO: 75; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 75; or any combination thereof.

[00699] In some embodiments, in a method of producing a steroidal alkaloid or steroidal saponin, said steroidal alkaloid or steroidal saponin comprises an anti-nutritive agent, a cosmetic agent, or a pharmaceutical agent or any combination thereof. In some embodiments, in a method of producing a steroidal alkaloid or steroidal saponin, said steroidal alkaloid or steroidal saponin is selected from an anti-nutritive agent, a cosmetic agent, or a pharmaceutical agent, or any combination thereof.

[00700] In some embodiments, in a method of producing a triterpenoid saponin, said triterpenoid saponin comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, or an anti-fungal agent, or any combination thereof. In some embodiments, in a method of producing a triterpenoid saponin, said triterpenoid saponin is selected from a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, or anti-fungal agent, or any combination thereof.

[00701] In some embodiments, in a method of producing a steroidal alkaloid, said steroidal alkaloid comprises alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, iun a method of producing a steroidal alkaloid, said steroidal alkaloid is selected from the group consisting of alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00702] In some embodiments, in a method of producing a steroidal saponin, said steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof. In some embodiments, iun a method of producing a steroidal alkaloid, said steroidal alkaloid is selected from the group consisting of uttroside B, a tomatoside, or any combination thereof.

[00703] In some embodiments, in a method of producing a triterpenoid saponin, said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, in a method of producing a triterpenoid saponin, said triterpenoid saponin is selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00704] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00705] In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a plant cell, a yeast cell, an alga cell, an insect cell, or a bacterium. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a plant cell, a yeast cell, or an alga cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a plant cell or a yeast cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a plant cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a yeast cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises an

alga cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises an insect cell. In some embodiments, in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, said cell comprises a bacterium.

[00706] In some embodiments, the at least one heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof.

[00707] In some embodiments, the step of introducing comprises transforming said at least one cell with said polynucleotide sequence or a vector comprising the polynucleotide sequence, and wherein said expression comprises transient expression or constituitive expression. In some embodiments, the vector comprises an expression vector. In some embodiments, the vector comprises a plasmid vector. In some embodiments, the vector integrates into the host cell DNA. In some embodiments, part of the polynucleotide comprised in the vector integrates into the host cell DNA. In some embodiments, the vector does not integrate into the host cell DNA and replicates as a separate entity within the host cell.

[00708] In some embodiments, expression vectors used to produce genetically modified plant cells include both *Agrobacterium* and non-*Agrobacterium* vectors. Agrobacterium-mediated gene transfer exploits the natural ability of Agrobacterium tumefaciens to transfer DNA into plant chromosomes and is described in detail in G. Gheysen, G. Angenon, and M. Van Montagu, 1998, *Agrobacterium*-mediated plant transformation: a scientifically intriguing story with significant applications in K. Lindsey (Ed.), Transgenic Plant Research, Harwood Academic Publishers, Amsterdam, pp. 1-33; and in H. A. Stafford (2000), Botanical Review 66:99-118. A second group of transformation methods is the non-*Agrobacterium*-mediated transformation and these methods are known as direct gene transfer methods. An overview is brought by P. Barcelo and P. A. Lazzeri (1998), Direct gene transfer: chemical, electrical and physical methods in K. Lindsey (Ed.), Transgenic Plant Research, Harwood Academic Publishers, Amsterdam, pp. 35-55. Methods include particle gun delivery, microinjection, electroporation of intact cells, polyethyleneglycol-mediated protoplast transformation, electroporation of protoplasts, liposome-mediated transformation, silicon-whiskers-mediated transformation, etc. A suitable control plant would include a genetically unaltered or non-transgenic plant of the parental line used to generate a transgenic plant herein.

[00709] Genetically transformed hairy root cultures can be obtained by transformation with virulent strains of *Agrobacterium rhizogenes*. Protocols used for establishing of hairy root cultures vary, as well as the susceptibility of plant species to infection by *Agrobacterium* (Toivounen et al. 1993; Vanhala et al. 1995). It is known that the *Agrobacterium* strain used for transformation has a great influence on root morphology and the degree of secondary metabolite accumulation in hairy

root cultures. It is possible by systematic clone selection, e.g., via protoplasts, to find high yielding, stable, and from single-cell-derived hairy root clones. This is possible because the hairy root cultures possess a great somaclonal variation. Another possibility of transformation is the use of viral vectors (Turpen 1999).

[00710] Any plant tissue or plant cells capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with an expression vector of interest. The team "organogenesis" means a process by which shoots and roots are developed sequentially from meristematic centers; the term "embryogenesis" means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include protoplasts, leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyls meristem).

[00711] In some embodiments, disclosed herein is a vector comprising a polynucleotide sequence or at least a heterologous gene encoding an enzyme as described in detail above. In some embodiments, disclosed herein is a host cell, for example but not limited to a plant cell, a yeast, or an algal cell, comprising a polynucleotide sequence or at least a heterologous gene encoding an enzyme as described in detail above, or a vector comprising a polynucleotide or at least one heterologous gene encoding an enzyme as described throughout.

[00712] The term "vector" in certain embodiments, encompasses a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been operably or continuously linked. The vector may be of any suitable type including, but not limited to, a phage, virus, plasmid, phagemid, cosmid, bacmid or even an artificial chromosome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., vectors having an origin of replication that functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, in some embodiments, vectors are capable of directing the expression of certain genes of interest. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). Suitable vectors have regulatory sequences, such as promoters, enhancers, terminator sequences, and the like as desired and according to a particular host organism (e.g., plant cell, yeast cell, or algal cell).

[00713] Typically, a recombinant vector according to the disclosure comprises at least one "heterologous gene encoding an enzyme" or "expression cassette comprising an at lest one

heterologous gene encoding an enzyme". Expression cassettes are generally DNA constructs, which in some embodiments include (5' to 3' in the direction of transcription): a promoter region, a polynucleotide sequence comprising at least one heterologous gene encoding an enzyme, homologue, variant, or fragment thereof of the disclosure as described above in detail, operably linked with the transcription initiation region, and a termination sequence including a stop signal for RNA polymerase and a polyadenylation signal. It is understood that all of these regions should be capable of operating in biological cells, such as plant cells or yeast cells or algal cells, to be transformed. The promoter region comprising the transcription initiation region, which preferably includes the RNA polymerase binding site, and the polyadenylation signal may be native to the biological cell to be transformed or may be derived from an alternative source, where the region is functional in the biological cell.

[00714] The term "recombinant host cell" ("expression host cell," "expression host system," "expression system" or simply "host cell"), as used herein, encompasses a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell that resides in a living tissue or organism. Host cells can be of fungal, plant or algal origin.

[00715] In some embodiments, disclosed herein is a transgenic plant or a cell derived thereof that is transformed with the above-described vector.

[00716] In some embodiments in methods of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the encoded enzyme comprises a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, or a cellulose synthase like G (CSLG). In some embodiments in methods of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, multiple enzymes are encoded by heterologous genes wherein the encoded enzymes comprises a combination of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, or a cellulose synthase like G (CSLG). In some embodiments in methods of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more enzymes are encoded by heterologous genes wherein the encoded enzymes comprises a combination of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, a UDP-glucose 6-dehydrogenase I, or a cellulose synthase like G (CSLG). In some embodiments in methods

of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the encoded enzyme comprises a CSLG.

[00717] In some embodiments, the cytochrome P450 comprises a C2-hydroxylase or is a C-23 oxidase. In some embodiments, the glycosyltransferase comprises a fructosyltransferase, a xylosyltransferase, or a UDP-glycosyltransferase, or a combination thereof. In some embodiments, the acyltransferase comprises a benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamyol/benzoyl-, deacetylvindoline (BAHD) acetyletransferase. In some embodiments, in a method of producing a triterpenoid saponin the encoded cytochrome P450 comprises a C2-hydroxylase or is a C-23 oxidase, or a combination thereof; said glycosyltransferase comprises a fructosyltransferase, a xylosyltransferase, or a UDP-glycosyltransferase, or a combination thereof; the said acyltransferase comprise a benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamyol/benzoyl-, deacetylvindoline (BAHD) acetyletransferase; or the encoded enzymes comprise a combination of any of these enzymes, wherein optionally additional enzymes are comprised within a polynucleotide sequence or sequences.

[00718] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the polynucleotide sequence of said at least one heterologous gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, wherein multiple heterologous genes are expressed, the skilled artisan would appreciate that a polynucleotide sequence or sequences may comprise a combination of any of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, as described in detail above. In some embodiments, wherein multiple heterologous genes are expressed, the skilled artisan would appreciate that a polynucleotide sequence or sequences may comprise a combination of comprises a nucleic acid sequence having at least 55% identity to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00719] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the polynucleotide sequence of said at least one heterologous gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or the polynucleotide sequence of said at least one heterologous gene comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth

in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, wherein multiple heterologous genes are expressed, the skilled artisan would appreciate that a polynucleotide sequence or sequences may comprise a combination of any of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105, as described in detail above. In some embodiments, wherein multiple heterologous genes are expressed, the skilled artisan would appreciate that a polynucleotide sequence or sequences may comprise a combination of comprises a nucleic acid sequence having at least 80% identity to the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 45, 46, 47, 51, 53, 55, 57, 59, 61, 63, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

[00720] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the polynucleotide sequence of said at least one heterologous gene encodes the amino acid sequence of said enzyme set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104; or the polynucleotide sequence of said at least one heterologous gene encodes the amino acid sequence of said enzyme set forth in the sequence having at least 55% identity to the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104.

[00721] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the polynucleotide sequence of said at least one heterologous gene encodes the amino acid sequence of said enzyme set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104; or the polynucleotide sequence of said at least one heterologous gene encodes the amino acid sequence of said enzyme set forth in the sequence having at least 80% identity to the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 48, 49, 50, 52, 54, 56, 58, 60, 62, 64, 64, 81, 94, 96, 98, 100, 102 or 104.

[00722] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell, for example but not limited to a plant cell, a yeast, or an algal cell comprises an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell. In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell, for example but not limited to a plant cell, a yeast, an algal cell, an insect cell, or a bacterium, comprises an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell. In some

embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell comprises a plant cell comprising an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell. In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell comprises a yeast cell comprising an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell. In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell comprises an algal cell comprising an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell. In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the genetically modified cell comprises a plant cell comprising an increased content of at least a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, compared to a corresponding unmodified cell, wherein said plant cell is comprised in a plant or plant part.

[00723] In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises an anti-nutritive agent, a cosmetic agent, or a pharmaceutical agent, or any combination thereof; or the steroidal alkaloid comprises an esculeoside or dehydroesculeoside; or the steroidal alkaloid comprises alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00724] In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-tomatine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises tomatine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises dehydrotomatine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-chaconine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-solanine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-solasonine. In some embodiments, in a method of producing a steroidal alkaloid, the steroidal alkaloid produced comprises alpha-solmargine.

[00725] In some embodiments, in a method of producing a steroidal saponin, the steroidal saponin produced comprises an anti-nutritive agent, a cosmetic agent, or a pharmaceutical agent, or any

combination thereof; or the steroidal alkaloid comprises uttroside B, a tomatoside, or any combination thereof. In some embodiments, in a method of producing a steroidal saponin, the steroidal saponin produced comprises uttroside B, a tomatoside, or any combination thereof.

[00726] In some embodiments, in a method of producing a steroidal saponin, the steroidal saponin produced comprises uttroside B. In some embodiments, in a method of producing a steroidal saponoin, the steroidal saponin produced comprises a tomatoside.

[00727] In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof; or the triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof;; or a combination thereof. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00728] In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises Compound 7. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises Compound 8. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises Compound 9. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin

produced comprises Compound 10. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises Compound 11 In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises glycyrrhizin (Compound 14). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises Glycyrrhetinic acid 3-O-monoglucuronide (compound 15). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises bayogenin (Compound 25). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises bayogenin-hexA-hex-hex (Compound 31). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises serjanic acid (Compound 26). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises serjanic acid-hexA-hex (Compound 32). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises soyasapogenol A (Compound 29). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises soyasapogenol B (Compound 30). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises soyasapogenol A–hexA-hex-pent (Compound 34). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises soyasaponin VI (Compound 35). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises betavulgaroside IV (Compound 33). In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises hederagenin-3GlcA, gypsogenin-3GlcA. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises gypsogenic acid-3GlcA. In some embodiments in a method of producing a triterpenoid saponin, the triterpenoid saponin produced comprises a QS-21 adjuvant, or any combination thereof.

[00729] In some embodiments, a method of producing a steroidal alkaloid, steroidal saponin, or triterpenoid saponin comprises increasing the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin, respectively, compared to an unmodified cell or an unmodified plant and altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00730] In some embodiments, a method of producing a steroidal alkaloid, steroidal saponin, or triterpenoid saponin comprises increasing the content of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin, respectively, compared to an unmodified cell or an unmodified plant and decreasing the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00731] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the cell into which the polynucleotide sequence is introduced comprises a plant cell or a yeast cell. Alternatively, in some embodiments, it comprises an algal cell. Plant and yeast cells that may be genetically modified as described herein, have been disclosed above in detail. Those same plant and yeast cells may, in some embodiments, be used in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin. In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the cell into which the polynucleotide sequence is introduced comprises a plant cell or a yeast cell or an algal cell or an insect cell or a bacterium.

[00732] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the cell comprises a plant cell that comprises a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a bract cell, a callus cell, and a flower cell.

[00733] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the yeast cell, the algal cell, or the plant cell is from an order, genus, or species recited herein. For example, but not limited to a plant cell comprising a cell from a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order; or a yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus.

[00734] In some embodiments of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the plant of the *Caryophyllales* order is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, and the *Rheum* genus; or the plant of the *Solanales* order is selected from the group of genera consisting of the *Nicotiana* genus, the *Solanum* genus, and the *Capsicum* genus; or the plant of the *Fabales* order is selected from the

group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Quillaja* genus, the *Glycine* genus, and the *Lotus* genus; or the plant of the *Apiales* order is selected from the group of genera consisting of the *Panax* genus, *Daucus* genus, the *Apium* genus, and the *Petroselinum* genus, or the plant from the *Poales* order is selected from the group of genera consisting of the *Oryza* genus (e.g., *Oryza sativa* and *Oryza glaberrima* [rice]), the *Hordeum* genus (e.g., *Hordeum vulgare* [barley]), the *Avena* genus (e.g., *Avena sativa* [oat], *Avena strigosa*), and the *Triticum* genus (e.g., *Triticum spelta* [spelt]), or the plant from the *Brassicales* order is selected from the group of genera consisting of the *Arabidopsis* genus (e.g., *Arabidopsis thaliana*), the *Brassica* genus (e.g., *Brassica oleracea* [cabbages], *Brassica juncea* [white mustard], *Brassica nigra* [black mustard], *Brassica napus*), the *Capparis* genus (e.g., *Capparis spinosa* [caper]), and the *Carica* genus (e.g., *Carica papaya* [papaya]).

[00735] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the plant of the *Solanales* order is selected from the group of species consisting of *Nicotiana benthamiana*, *Solanum lycopersicum, Solanum tuberosum, Solanum melongena, Solanum pennellii, Solanum chacoense*, *Solanum dulcamara,* and *Capsicum annuum*; or the plant of the *Fabales* order is selected from the group of species consisting of *Glycyrrhiza uralensis, Medicago sativa, Medicago truncatula, Quillaja saponaria, Glycine max*, and *Lotus japonicus*; or the plant of the *Malvales* order is selected from the *Theobroma* genus; or the plant of the *Apiales* order is selected from the group of species consisting of *Panax ginseng, Daucus carota, Apium graveolens*, and *Petroselinum crispum*; or the plant is selected from the species *Theobroma cacao,* or the plant from the *Poales* order is selected from the group of genera consisting of the *Oryza* genus (e.g., *Oryza sativa* and *Oryza glaberrima* [rice]), the *Hordeum* genus (e.g., *Hordeum vulgare* [barley]), the *Avena* genus (e.g., *Avena sativa* [oat], *Avena strigosa*), and the *Triticum* genus (e.g., *Triticum spelta* [spelt]), or the plant from the *Brassicales* order is selected from the group of genera consisting of the *Arabidopsis* genus (e.g., *Arabidopsis thaliana*), the *Brassica* genus (e.g., *Brassica oleracea* [cabbages], *Brassica juncea* [white mustard], *Brassica nigra* [black mustard], *Brassica napus*), the *Capparis* genus (e.g., *Capparis spinosa* [caper]), and the *Carica* genus (e.g., *Carica papaya* [papaya]).

[00736] In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus. In some embodiments, said yeast is a *Saccharomyces cerevisiae* or *Candida albicans.* In some embodiments in a method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the method further comprises a step of purifying said steroidal alkaloid, steroidal saponin, or triterpenoid

saponin.

[00737] Methods of purifying steroidal alkaloids, steroidal saponins, or triterpenoids may include standard purifications techniques known in the art, for example but not limited to column liquid chromatography, preparative medium-pressure liquid chromatograph, and preparative HPLC. These techniques may be used in embodiments of purifying steroidal alkaloids, steroidal saponins, or triterpenoid saponins from plants, plant part, or plant cells. Alternatively, the plant cells or plant parts may be used as is without purifying the steroidal alkaloid, steroidal saponin, or triterpenoid saponin.

[00738] In some embodiments, when steroidal alkaloids, steroidal saponins, or triterpenoid saponins are produced in yeast cells, the steroidal alkaloids, steroidal saponins, or triterpenoid saponins are secreted from the cells. In some embodiments, a secreted steroidal alkaloid, steroidal saponin, or triterpenoid saponin is purified from a yeast cell medium.

[00739] In some embodiments, a steroidal alkaloid, steroidal saponin or triterpenoid saponin is extracted from a plant cell, a plant part, a plant, algae, a yeast, an insect cell, or bacterial culture using methods known in the art.

[00740] In some embodiments, provided herein is a method of producing at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin, the method further comprising the step of extracting said steroidal alkaloid, steroidal saponin, or triterpenoid saponin from the plant, plant part, colony, organ, tissue, or cells that produce said steroidal alkaloid, steroidal saponin, or triterpenoid saponin. In one embodiment, the plant part comprises a leaf , a petiole, a plant stem or stalk, a root, a bud, a tuber, a bean, a grain or kernel, a fruit, a nut, a legume, a seed or seed, a bract, or a flower.

Methods of Altering the Content of Steroidal Alkaloids, Steroidal Saponins, and Triterpenoid Saponins in a Plant or Plant Part or Plant Cell

[00741] In some embodiments, disclosed herein is a method of altering the content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a genetically modified plant, a genetically modified plant part, or a genetically modified plant cell, the method comprising genetically modifying the CSLG gene in said plant, plant part, or said cell, wherein said modification comprises increasing expression of said CSLG gene, or increasing activity of the expressed CSLG enzyme, or increasing stability of the expressed CSLG enzyme, or any combination thereof; wherein said plant, plant part, or plant cell comprises an increased content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin compared to a corresponding unmodified plant, plant part, or plant cell. In other embodiments, disclosed herein is a method of altering the content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a genetically modified plant, a genetically modified plant part, or a genetically modified plant cell, the method comprising

genetically modifying the CSLG gene in said plant, plant part, or said cell, wherein said modification comprises decreasing or eliminating expression of said CSLG gene, or decreasing activity of the expressed CSLG enzyme, or decreasing stability of the expressed CSLG enzyme, or any combination thereof; wherein said plant, plant part, or cell comprises a decreased content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin compared to a corresponding unmodified plant, plant part, or plant cell.

[00742] In some embodiments, a genetically modified plant described in detail herein, comprises an altered content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin disclosed herein.

[00743] In certain embodiments, disclosed herein is a method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or a plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising:

(a) transforming said at least one plant cell with at least one silencing molecule targeted to a nucleic acid gene sequence encoding a Cellulose Synthase Like G (CSLG) enzyme; or

(b) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence;

wherein expression of the gene encoding the CSLG enzyme is reduced in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell, wherein the plant comprising said genetically modified cell comprises reduced content of one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00744] In certain embodiments, disclosed herein is a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic

intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising:

(a) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; and

(b) expressing said nucleic acid encoding said CSLG; and

wherein the plant comprising said genetically modified cell comprises increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

[00745] Methods for mutagenizing at least one nucleic acid sequence encoding a steroidal alkaloid biosynthetic enzyme, a steroidal saponin biosynthetic enzyme, or a triterpenoid saponin biosynthetic enzyme, for example but not limited to a CSLG enzyme, have been described in detail above under the header *Genetically Modified Cells & Genetically Modified Plants.* Similarly, methods for transforming a plant cell with at least one silencing molecule targeted to a nucleic acid sequence encoding a steroidal alkaloid biosynthetic enzyme, a steroidal saponin biosynthetic enzyme, or a triterpenoid saponin biosynthetic enzyme, for example a nucleic acid sequence encoding a CSLG enzyme, have been described in detail above under the header *Genetically Modified Cells & Genetically Modified Plants.* These methods are incorporated herein in their entirety, wherein a skilled artisan would appreciate that the methods used to produce a genetically modified cell comprised within a genetically modified plant or plant part are in certain embodiments, the same methods used to reduce or increase the content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a plant or plant part, especially as the genetically modified plants or plant parts described have an altered content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof.

[00746] In some embodiments, when the at least one silencing molecule is targeted to a nucleic acid gene sequence encoding a CSLG enzyme, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36 38, 40, 65, 80, 93, 95, 97, 99, 101,

103, or 105. In some embodiments, when a nucleic acid gene sequence encoding a CSLG enzyme is mutated, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105. In some embodiments, in a method of increasing or a method of decreasing an at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin, the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102 or 104.

[00747] In some embodiments, derivatives of steroidal alkaloids, steroidal saponins, or triterpenoid saponins comprise glycosylated derivatives of, respectively, steroidal alkaloids, steroidal saponins, or triterpenoid saponins.

[00748] In some embodiments, the mutation comprises a mutation in a non-coding region. In some embodiments, the mutation comprises a mutation in a coding region. Not limiting examples of mutations in non-coding regions are those mutations that increase or decrease the expression of the CSLG gene. In some embodiments, the mutation comprises overexpression of the CSLG gene, wherein the genetically modified plant cell comprises an increase in at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. Methods for overexpression are described in detail herein wherein the skilled artisan would appreciate that a constitutive or inducible promoter may be incorporated into the construct comprising the nucleic acid sequence encoding the CSLG gene. In some embodiments, the mutation comprises reduced expression of the CSLG gene, wherein the genetically modified plant cell comprises a decrease in at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin. Methods for specifically targeting a nucleic acid sequence in order to reduce expression are detailed throughout, wherein the skilled artisan would appreciate that reduction of expression would in certain embodiments, lead to reduction in content of at least one steroidal alkaloid, steroidal saponin, or triterpenoid saponin.

[00749] In some embodiments, in a method of increasing the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof of the gene encoding the CSLG enzyme is increased in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell; or said encoded CSLG enzyme has increased activity in the genetically modified plant cell compared to its activity in a corresponding unmodified

plant cell; or said encoded CSLG enzyme has increased stability in the genetically modified plant cell compared to its stability in a corresponding unmodified plant cell; or any combination thereof. Methods for mutation of nucleic acids encoding genes of the steroidal alkaloid, steroidal saponin, or triterpenoid saponin biosynthetic pathway, for example but not limited to the CSLG gene, are described in detail above and include mutation the coding region for increased stability or activity or a combination thereof of the enzyme, and or mutation the nucleic acid sequence expressing the CSLG gene for over expression. One skilled in the art would appreciate that those methods of mutation could in certain embodiments be used herein as well.

[00750] In some embodiments, a method of reducing at least one steroidal alkaloid reduces a steroidal alkaloid comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal alkaloid in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one steroidal alkaloid eliminates a steroidal alkaloid comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal alkaloid in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one steroidal alkaloid reduces a steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, eliminates a steroidal alkaloid, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal alkaloid in a plant or plant part comprising a non-modified plant cell.

[00751] In some embodiments, the at least one steroidal alkaloid comprises an esculeoside or a dehydroesculeoside. In some embodiments, the at least one steroidal alkaloid comprises alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, the reduced at least one steroidal alkaloid comprises an esculeoside or a dehydroesculeoside. In some embodiments, the reduced at least one steroidal alkaloid comprises alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, the reduced at least one steroidal alkaloid biosynthetic intermediate comprises cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26 amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, or beta-1-tomatine, or any combination thereof.

[00752] In some embodiments, the reduced at least one steroidal alkaloid comprises eliminating or nearly eliminating an esculeoside or a dehydroesculeoside. In some embodiments, the reduced at least one steroidal alkaloid comprises eliminating or nearly eliminating alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof. In some embodiments, the reduced at least one steroidal alkaloid biosynthetic intermediate comprises eliminating or nearly eliminating cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26 amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, or beta-1-tomatine, or any combination thereof. In some embodiments, a method of reducing at least one steroidal alkaloid comprises eliminating a combination of at least one steroidal alkaloid and at least one steroidal alkaloid intermediate.

[00753] In some embodiments, a method of increasing at least one steroidal alkaloid increases a steroidal alkaloid comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal alkaloid in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of increasing the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a steroidal alkaloid comprising an esculeoside or a dehydroesculeoside. In some embodiments, a method of increasing the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a steroidal alkaloid comprising alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00754] In some embodiments, a method of increasing the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases at least one steroidal alkaloid biosynthetic intermediate comprises cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone, furostanol-26-aldehyde, 26-amino-furostanol, tomatidenol, tomatidine, tomatidine galactoside, gamma-tomatine, or beta-1-tomatine, or any combination thereof. In some embodiments, a method of increasing the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a combination thereof.

[00755] In some embodiments, a method of reducing at least one steroidal saponin reduces a steroidal saponin comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal saponin in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one steroidal saponin eliminates a steroidal saponin comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal saponin in a plant or plant part comprising a non-modified plant cell. In some

embodiments, a method of reducing at least one steroidal saponin reduces a steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, eliminates a steroidal saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal saponin in a plant or plant part comprising a non-modified plant cell.

[00756] In some embodiments, the at least one steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof. In some embodiments, the reduced at least one steroidal saponin comprises uttroside B, a tomatosideor any combination thereof. In some embodiments, the reduced at least one steroidal alkaloid biosynthetic intermediate comprises cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone.

[00757] In some embodiments, the reduced at least one steroidal saponin comprises eliminating or nearly eliminating uttroside B, a tomatosideor any combination thereof. In some embodiments, the reduced at least one steroidal alkaloid biosynthetic intermediate comprises eliminating or nearly eliminating cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone. In some embodiments, a method of reducing at least one steroidal saponin comprises eliminating a combination of at least one steroidal saponin and at least one steroidal saponin biosynthetic intermediate.

[00758] In some embodiments, a method of increasing at least one steroidal saponin increases a steroidal saponin comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the steroidal saponin in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of increasing the content of a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a steroidal saponin comprising uttroside B, a tomatoside, or any combination thereof.

[00759] In some embodiments, a method of increasing the content of a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases at least one steroidal saponin biosynthetic intermediate comprising cholesterol, 22-hydroxycholesterol, 22,26-dihydroxycholesterol, furostanol-type saponin aglycone or any combination thereof. In some embodiments, a method of increasing the content of a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a combination thereof.

[00760] In some embodiments, a method of reducing at least one triterpenoid saponin reduces a

triterpenoid saponin comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the triterpenoid saponin in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one triterpenoid saponin eliminates a triterpenoid saponin comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the triterpenoid saponin in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one triterpenoid saponin reduces a triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, in a plant or plant part comprising a non-modified plant cell. In some embodiments, a method of reducing at least one triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, eliminates a triterpenoid saponin, derivative thereof, metabolite thereof, or biosynthetic intermediate thereof, comprising a toxin or a bitter tasting compound or a combination thereof, compared to the content of the triterpenoid saponin in a plant or plant part comprising a non-modified plant cell.

[00761] In some embodiments, the reduced at least one triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, the reduced triterpenoid saponin intermediate comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a method of reducing at least one triterpenoid saponin comprises reducing a combination of at least one triterpenoid saponin and one triterpenoid saponin biosynthetic intermediate.

[00762] In some embodiments, the reduced at least one triterpenoid saponin comprises eliminating or nearly eliminating medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound

32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, the reduced triterpenoid saponin intermediate comprises eliminating or nearly eliminating Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a method of reducing at least one triterpenoid saponin comprises eliminating a combination of at least one triterpenoid saponin and one triterpenoid saponin biosynthetic intermediate.

[00763] In some embodiments, a method of increasing the content of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a triterpenoid saponin comprising a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof. In some embodiments, a method of increasing the content of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a triterpenoid saponin comprising medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof. In some embodiments, a method of increasing the content of a triterpenoid saponin, increases the triterpenoid saponin glycyrrhizin (Compound 14). In some embodiments, a method of increasing the content of a triterpenoid saponin, increases the triterpenoid saponin QS-21 adjuvant.

[00764] In some embodiments, a method of increasing the content of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a triterpenoid saponin biosynthetic intermediate comprising Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof. In some embodiments, a method of increasing the content of a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof increases a combination thereof.

[00765] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, increases the content of the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and alters the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00766] In some embodiments, a method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, increases the content of the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, and decreases the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00767] In some embodiments, the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, decreases the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a

biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00768] In some embodiments, the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, decreases the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof and an increased content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

[00769] In some embodiments in a method of altering (increasing or decreasing) the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof. the method comprises genetically modifying at least one plant cell, wherein in some embodiments said plant cell comprises a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a bract cell, a callus cell, and a flower cell.

[00770] In some embodiments in a method of altering (increasing or decreasing) the content of a steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof. the method comprises genetically modifying at least one plant cell, wherein in some embodiments said plant cell comprises an order, genus, or species recited herein in detail. For

example, but not limited to a plant cell comprising a cell from a plant in the *Poales* order, the Caryophyllales order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

[00771] In some embodiments of methods altering the content (increasing or decreasing) of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the plant of the *Caryophyllales* order is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, and the *Rheum* genus; or the plant of the *Solanales* order is selected from the group of genera consisting of the *Nicotiana* genus, the *Solanum* genus, and the *Capsicum* genus; or the plant of the *Fabales* order is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Quillaja* genus, the *Glycine* genus, and the *Lotus* genus; or the plant of the *Apiales* order is selected from the group of genera consisting of the *Panax* genus, *Daucus* genus, the *Apium* genus, and the *Petroselinum* genus, or the plant from the *Poales* order is selected from the group of genera consisting of the *Oryza* genus (e.g., *Oryza sativa* and *Oryza glaberrima* [rice]), the *Hordeum* genus (e.g., *Hordeum vulgare* [barley]), the *Avena* genus (e.g., *Avena sativa* [oat], *Avena strigosa*), and the *Triticum* genus (e.g., *Triticum spelta* [spelt]), or the plant from the *Brassicales* order is selected from the group of genera consisting of the *Arabidopsis* genus (e.g., *Arabidopsis thaliana*), the *Brassica* genus (e.g., *Brassica oleracea* [cabbages], *Brassica juncea* [white mustard], *Brassica nigra* [black mustard], *Brassica napus*), the *Capparis* genus (e.g., *Capparis spinosa* [caper]), and the *Carica* genus (e.g., *Carica papaya* [papaya]).

[00772] In some embodiments of methods altering the content (increasing or decreasing) of a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin, the plant of the *Solanales* order is selected from the group of species consisting of *Nicotiana benthamiana*, *Solanum lycopersicum, Solanum tuberosum, Solanum melongena, Solanum pennellii, Solanum chacoense*, *Solanum dulcamara,* and *Capsicum annuum*; or the plant of the *Fabales* order is selected from the group of species consisting of *Glycyrrhiza uralensis, Medicago sativa, Medicago truncatula, Quillaja saponaria, Glycine max*, and *Lotus japonicus*; or the plant of the *Malvales* order is selected from the *Theobroma* genus; or the plant of the *Apiales* order is selected from the group of species consisting of *Panax ginseng, Daucus carota, Apium graveolens*, and *Petroselinum crispum*; or the plant is selected from the species *Theobroma cacao*.

Methods of Producing a Steroidal Alkaloid, Steroidal Saponin, or Triterpenoid Saponin in an In Vitro System

[00773] *In vitro* systems are well known in the art. In some embodiments, an *in vitro* translation system used in the methods for producing a triterpenoid saponin disclosed herein comprises a rabbit reticulocyte lysate, a wheat germ extract, or an *E. coli* cell-free system. Detailed methods for using a

cell free system and known and freely available in the art, for example but not limited to in formation found at: https://www.thermofisher.com/il/en/home/references/ambion-tech-support/large-scale-transcription/general-articles/the-basics-in-vitro-translation.html.

[00774] In some embodiments, disclosed herein is a method of producing a steroidal alkaloid or a steroidal saponin in an *in vitro* system, the method comprising:

(a) combining a nucleic acid sequence or nucleic acid sequences encoding a cytochrome P450 (*GAME7*) gene, a cytochrome P450 (*GAME8*) gene, a 2-oxoglutarate-dependent dioxygenase (*GAME11*) gene, a cytochrome P450 (*GAME6*) gene, a cytochrome P450 (*GAME4*) gene, a transaminase (*GAME 12*) gene, a tomatidine UDP-galactosyltransferase (*GAME1/SGT1*) gene, a UDP-gycosyltransferase (*GAME17*) gene, a UDP-gycosyltransferase (*GAME18*) gene, a UDP-gycosyltransferase/xylosyltransferase (*GAME2/SGT3*) gene, and a cellulose synthase like G gene (CSLG or *GAME15*), wherein said nucleic acid sequence or nucleic acid sequences are optionally comprised in a vector or vectors; and

(b) expressing said genes;

(c) incubating a combination of said expressed enzymes together;

wherein the product of said incubation comprises at least one steroidal alkaloid or steroidal saponin.

[00775] In some embodiments, the above nucleic acid sequence(s) further comprise an ethylene-responsive element binding factor 13 (*GAME9*) and/or a BHLH-transcription factor.

[00776] In some embodiments, disclosed herein is a method of producing a steroidal alkaloid or a steroidal saponin in an *in vitro* system, the method comprising:

(d) combining a nucleic acid sequence or nucleic acid sequences encoding a cytochrome P450 (*GAME7*) gene, a cytochrome P450 (*GAME8*) gene, a 2-oxoglutarate-dependent dioxygenase (*GAME11*) gene, a cytochrome P450 (*GAME6*) gene, and a cellulose synthase like G gene (CSLG or *GAME15*), wherein said nucleic acid sequence or nucleic acid sequences are optionally comprised in a vector or vectors; and

(e) expressing said genes;

(f) incubating a combination of said expressed enzymes together;

wherein the product of said incubation comprises at least one steroidal alkaloid or a steroidal saponin, or a combination thereof. See for example but not limited to, the biosynthetic pathways and resultant inermediates and products presented therein in **Figures 1, 8D, 13,** and **14A-14D**.

[00777] In some embodiments, the biosynthetic GAME genes and/or the CSLG gene are operably linked to a promoter, a transcription termination sequence, or a combination thereof as has been described in detail herein for making constructs. In some embodiments, in a method of producing a

steroidal alkaloid or a steroidal saponin, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101. In some embodiments, in a method of producing a steroidal alkaloid or a steroidal saponin, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101.

[00778] In some embodiments, the biosynthetic GAME genes and/or the CSLG gene are operably linked to a promoter, a transcription termination sequence, or a combination thereof as has been described in detail herein for making constructs. In some embodiments, in a method of producing a steroidal alkaloid or a steroidal saponin, the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101.

[00779] In certain embodiments, an in vitro method of producing a steroidal alkaloid comprises producing a steroidal alkaloid selected from a esculeoside or a dehydroesculeoside. In certain embodiments, an *in vitro* method of producing a steroidal alkaloid comprises producing a steroidal alkaloid selected from alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof.

[00780] In certain embodiments, an *in vitro* method of producing a steroidal saponin comprises producing a steroidal saponin selected from uttroside B, a tomatosideor any combination thereof.

[00781] In some embodiments, disclosed herein is a method of producing a triterpenoid saponin in an *in vitro* system, the method comprising:

(g) combining a nucleic acid sequence or nucleic acid sequences encoding a saponin beta-amyrin synthase gene (SOAP1), a cytochrome P450 gene (SOAP2), a cytochrome P450 gene expressing a C-2 hydroxylase (SOAP3), a cytochrome P450 gene expressing a C-23 oxidase (SOAP4), a glycosyl transferase gene expressing a UDP-glycosyltransferase or a fucocyl transferase (SOAP6), a glycosyl transferase gene expressing a UDP-glycosyltransferase (SOAP7), a glycosyl transferase gene expressing a UDP-glycosyltransferase (SOAP8); a glycosyl transferase gene expressing a UDP-glycosyltransferase or a xylosyl transferase (SOAP9), an acyltransferase gene (SOAP10), and a cellulose synthase like G gene (CSLG), wherein said nucleic acid sequence or nucleic acid sequences are optionally comprised in a

vector or vectors; and

(h) expressing said genes;

(i) incubating a combination of said expressed enzymes together;

wherein the product of said incubation comprises at least one triterpenoid saponin.

[00782] In some embodiments, the biosynthetic SOAP genes and/or the CSLG gene are operably linked to a promoter, a transcription termination sequence, or a combination thereof as has been described in detail herein for making constructs. In some embodiments, in a method of producing a triterpenoid saponin the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101. In some embodiments, in a method of producing a triterpenoid saponin the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101.

[00783] In some embodiments, the biosynthetic SOAP genes and/or the CSLG gene are operably linked to a promoter, a transcription termination sequence, or a combination thereof as has been described in detail herein for making constructs. In some embodiments, in a method of producing a triterpenoid saponin the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101.

[00784] In some embodiments, the nucleic acid gene sequence or sequences encoding the other triterpenoid saponin biosynthetic enzymes are set forth as follows:

(a) the nucleic acid sequence encoding said β-amyrin synthase gene is set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; and

(b) the nucleic acid sequence or sequences encoding said cytochrome P450 genes, are set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; and

(c) the nucleic acid sequence or sequences encoding said glycosyl transferase genes, are set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of

SEQ ID NO: 55, 57, 59, or 61; and

(d) the nucleic acid sequence encoding said acyltransferase is set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63.

[00785] In certain embodiments, an *in vitro* method of producing a triterpenoid saponin comprises producing a triterpenoid saponin selected from a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof.

[00786] In certain embodiments, an *in vitro* method of producing a triterpenoid saponin comprises producing a triterpenoid saponin selected from medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof.

[00787] As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a molecule" also includes a plurality of molecules.

[00788] Throughout this application, various embodiments may be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub ranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed sub ranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[00789] Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases “ranging/ranges between” a first indicate number and a second indicate number and “ranging/ranges from” a first indicate number “to” a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals there between.

[00790] Further, one skilled in the art would appreciate that the term "comprising" used throughout is intended to mean that the genetically modified or gene edited plants disclosed herein, and methods of altering expression of genes, and altering production of SA and/or SGA within these genetically modified or gene edited plants includes the recited elements, but not excluding others which may be optional. "Consisting of" shall thus mean excluding more than traces of other elements. The skilled artisan would appreciate that while, in some embodiments the term "comprising" is used, such a term may be replaced by the term "consisting of", wherein such a replacement would narrow the scope of inclusion of elements not specifically recited.

[00791] The following examples are presented in order to more fully illustrate some embodiments of the invention. They should, in no way be construed, however, as limiting the broad scope of the invention. They should, in no way be construed, however, as limiting the broad scope of the genetically modified cells and plants disclosed herein or the uses of said cells and plants to produce steroidal alkaloids, steroidal saponins, or triterpenoid saponins; or methods for increasing or decreasing the content of a steroidal alkaloid, steroidal saponin, or triterpenoid saponin in a plant or plant part. One skilled in the art can readily devise many variations and modifications of the principles disclosed herein without departing from the scope disclosed herein.

EXAMPLES

Example 1: Materials and Methods for Examples 2-13

Plant material, treatments, and generation of transgenic plants

[00792] Tomato (*Solanum lycopersicum*; cv. Micro Tom) and potato (*Solanum tuberosum*; cultivar Desiree) plants were collected as described previously (Itkin et al., 2001, *supra*). In potato, when the green parts started to dry, mature tubers (Stage 3) were collected, washed of soil, dried and kept at 4°C, at complete darkness.

[00793] The GAME9-silenced (RNAi) and overexpression (OX) constructs were created by introducing the corresponding GAME9 DNA fragments to pK7GWIWG2(II) and pJCV52 binary vectors, respectively. Transgenic lines for silencing and overexpression of GAME9 in tomato and potato were generated and tissue extracts were prepared and analyzed according to Itkin et al. (2011, *supra*).

[00794] **Table 3** below describes the oligonucleotides used for generation of the constructs described herein. The *GAME4*-silencing (RNAi; *GAME4*i), *GAME4* overexpressing (*GAME4*oe) and *GAME8*-silencing constructs were generated as described previously (Itkin et al., 2001, *supra*; WO 2012/095843).

[00795] **Table 3: Oligonucleotides used for construct production**

Name	**Sequence 5` to 3`/ Description**	SEQ ID NO.
Sl07g0434 20 EcoRI Fw	AAAAAgaattcCGGATCTTCTCTCGAACTGGTCAA To prepare *GAME11* virus-induced gene silencing (VIGS) construct	1
Sl07g0434 20 EcoRI Rv	AAAAAgaattcCACTTTCATTGCTTCATCCATTAGATCT To prepare *GAME11* VIGS construct	2
Sl07g0435 00 EcoRI Fw	AAAAAgaattcCTTAGCTTATGGCCACATCACACCTT To prepare *GAME18* VIGS construct	3
Sl07g043500 EcoRI Rv	AAAAAgaattcACTCAAGATTTGGTGAAGCTGTGGTT To prepare *GAME18* VIGS construct	4
G8-Forward (AscI)	AAAAAGGCGCGCCAATCATAGAGAAGAAAGAAGACG To construct RNAi of *GAME8*	5
G8-Reverse (Not I)	AAAAAGCGGCCGCACTCCTGCAGGAATTGTCATTTCTC To construct RNAi of *GAME8*	6
GAME9 RNAi NotI Fw	aaaaaGCGGCCGCATGAGTATTGTAATTGATGATGATGAAATC To construct RNAi of *GAME9*	7
GAME9 RNAi AscI Rv	aaaaGGCGCGCCCACACGCCACAGATGGTTCTT To construct RNAi of *GAME9*	8
GAME9-Tom GW Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAGTATTGTAATTGATGATGATGAAATC To pick up the gene from cDNA for overexpression (good for tomato)	9
GAME9-Tom GW Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATACTACCTTCTGTCCTAAGCCT To pick up the gene from cDNA for overexpression (good for tomato)	10
GAME9-Pot GW Fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAATATTGCAATTGATGATGATGA To pick up the gene from cDNA for overexpression (good for potato)	11
GAME9-Pot GW Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATTTGTATCAACATTTGTAAATTCACAC To pick up the gene from cDNA for overexpression (good for potato)	12

Co-expression analysis

[00796] The tomato GAME1 (Solyc07g043490) and its potato ortholog SGT1 (PGSC003DMG400011749) were used as 'baits' in the co-expression analysis, resulting in lists (sorted in descending order by *r*-value $\geq$ 0.8) of co-expressed genes (for each 'bait' separately). Two

homologous genes were subsequently identified (Solyc12g006460 and PGSC0003DMG400024274 in tomato and potato, respectively), which were highly correlated with the "bait" genes (*r*-value > 0.9 in both species). Those genes were identified as *GLYCOALKALOID METABOLISM 4* (*GAME4*, WO 2012/095843). The *GAME4* genes were further added as 'baits' to the previous (*GAME1*) co-expression analysis. The co-expression lists for *GAME1* (*SGT1*) and *GAME4* in both species were used to construct co-expression correlation network. The analysis was performed as follows: tomato RNAseq transcriptome data from different tissues and organs (flesh, peel, seeds, roots, leaves, buds, flowers, pollen) and developmental stages (25 experiments in total) (Itkin et al., 2011, *ibid*) and potato RNAseq transcriptome data from different tissues and organs (40 experiments in total) (US 2012/0159676), were used. First, an R script was used to perform co-expression analysis (for each species) and the list of co-expressed genes was constructed as a FASTA file, using a Perl script. Finally, BLASTall tools (Camacho C. et al., 2009. BMC Bioinform 10:421) were used to find shared homologs between the two species. The tblastx criteria for homolog similarity were set to *p*-value > 0.05, minimum 25 nucleotides, and at least 60 percent similarity as an overall identity for each gene. The co-expression network was visualized with the Cytoscape program (Shannon P. et al., 2003. Genome Res. 13:2498–2504).

Phylogenetic analysis

[00797] The protein sequences were aligned using the Muscle algorithm and the phylogenetic tree was analyzed and visualized by the SeaView v4.3.5 program using the maximum likelihood method by PhyML 3.0 (Expósito-Rodríguez M et al., 2008. BMC Plant Biol. 8:131) with the following settings: model - LG; The approximate likelihood ratio test (aLRT) Shimodaira-Hasegawa-like (SH-like) procedure was used as a statistical test to calculate branch support (branch support – aLRT (SH-like)); invariable sites - optimized; across site rate variation - optimized; tree searching operations – best for NNI & SPR; starting tree – BioNJ, optimize tree topology. The numbers on the branches indicate the fraction of bootstrap iterations supporting each node. The accession numbers of the proteins used for the preparation of this tree and the organism names are listed in **Table 4** hereinbelow; the tree is presented in **FIGURE 12**.

[00798] **Table 4: Accession numbers of the sequences used for the construction of the phylogenetic tree**

Name as appears in FIGURE 12	***Latin and common name***	**Accession number**
GuCYP88D6	*Glycyrrhiza uralensis*	BAG68929.1
LjCYP88D4	*Lotus japonicus*	BAG68927.1
MtCYP88D3	*Medicago truncatula*	BAG68926.1
CmCYP88A2	*Cucurbita maxima*	AF212991
AtCYP88A3	*Arabidopsis thaliana*	AAB71462.1

Name as appears in FIGURE 12	***Latin and common name***	**Accession number**
PsCYP88A7	*Pisum sativum*	AAO23064.1
ZmCYP88A1	*Zea mays*	NP_001105586.1
GmCYP88A26	*Glycine max*	XP_003516638.1
CaCYP89A35	*Capsicum annuum*	DQ114394
GmCYP89A36	*Glycine max*	DQ340245
ZmCYP89B17	*Zea mays*	CO465851.1
TmCYP89J1	*Triticum monococcum*	AY914081
SlCYP88B1 (GAME4)	*Solanum lycopersicum*	Solyc12g006460.1.1
SpimpCYP88B1 (GAME4)	*Solanum pimpinellifolium*	contig 6356779
SpCYP88B1 (GAME4)	*Solanum pinelii*	AW618484.1, BG135958.1
StCYP88B2 (GAME4)	*Solanum tuberosum* group *Phureja*	PGSC0003DMP400041994
StCYP88B1v2 (GAME4)	*Solanum tuberosum* group *Tuberosum*	PGSC0003DMP400041994
SlCYP88C2	*Solanum lycopersicum*	Solyc10g007860.2.1
SmCYP88B3 (GAME4)	*Solanum melongena*	FS071104, FS071103
OsCYP90A3	*Oryza sativa*	AC123526.1
SlCYP90A5	*Solanum lycopersicum*	Solyc06g051750.2.1
ScCYP90A8	*Citrus sinensis*	DQ001728.1
ZeCYP90A11	*Zinnia elegans*	BAE16977.1
PhCYP88C1	*Petunia hybrida*	AAZ39647.1
AaCYP90A13	*Artemisia annua*	ABC94481.1
AtCYP710A1	*Arabidopsis thaliana*	AAC26690.1
SmCYP71A2	*Solanum melongena*	X71654.1
GmCYP93E1	*Glycine max*	AB231332
HlCYP71C25	*Hordeum lechleri*	AY462228
NtCYP71D16	*Nicotiana tabacum*	AF166332
MeCYP71E7	*Manihot esculenta*	AY217351
TaCYP71F1	*Triticum aestivum*	AB036772
AoCYP71J1	*Asparagus officinalis*	AB052131
MaCYP71N1v2	*Musa acuminata*	AY062167
TaCYP72A6v1	*Triticum aestivum*	AF123604
ZmCYP72A16	*Zea mays*	AF465265
LeCYP72A51	*Solanum lycopersicum*	Solyc10g051020.1.1
GmCYP72A61	*Glycine max*	DQ340241
MtCYP716A12	*Medicago truncatula*	ABC59076.1
StCYP716A13	*Solanum tuberosum*	PGSC0003DMP400013378
AaCYP716A14	*Artemisia annua*	DQ363134
PsCYP716B2	*Picea sitchensis*	AY779543
SlCYP718A6	*Solanum lycopersicum*	Solyc07g055970.1.1
MtCYP718A8	*Medicago truncatula*	XP_003617455.1
PsCYP719B1	*Papaver somniferum*	EF451150
StCYP72A186 (GAME7)	*Solanum tuberosum*	PGSC0003DMG402012386
SlCYP72A186 (GAME7)	*Solanum lycopersicum*	Solyc07g062520
SlCYP72A188 (GAME6)	*Solanum lycopersicum*	Solyc07g043460
StCYP72A188 (GAME6)	*Solanum tuberosum*	PGSC0003DMG400011750

Name as appears in FIGURE 12	*Latin and common name*	Accession number
GuCYP72A154	*Glycyrrhiza uralensis*	BAL45206.1
MtCYP72A59	*Medicago truncatula*	ABC59078.1
NtCYP72A57	*Nicotiana tabacum*	ABC69414.1
NtCYP72A54	*Nicotiana tabacum*	ABC69417.1
CrCYP72A1	*Catharanthus roseus*	gi461812
MtCYP72A63	*Medicago truncatula*	gi371940452
NpCYP72A2	*Nicotiana plumbaginifolia*	AAB05376.3
SlCYP734A7	*Solanum lycopersicum*	Solyc03g120060.1.1
StCYP72A29	*Solanum tuberosum*	BAB86912.1
StSYP72a56	*Solanum tuberosum*	PGSC0003DMG400017325
StCYP72A208 (GAME8a)	*Solanum tuberosum*	PGSC0003DMG400026594
StCYP72A208 (GAME8b)	*Solanum tuberosum*	PGSC0003DMG400026586
SlCYP72A208 (GAME8a)	*Solanum lycopersicum*	TC243022
SlCYP72A208 (GAME8b)	*Solanum lycopersicum*	SGN-U578058

Metabolite analysis

[00799] Preparation of plant tissue extracts and profiling of semi-polar compounds (including steroidal alkaloids and steroidal saponins) by UPLC–qTOF–MS and phytosterol content of the tomato leaves were carried out as described previously (Itkin et al., 2011, *supra*).

Quantitative Real-Time PCR assays

[00800] RNA was isolated and Quantitative Real-Time PCR was performed as described previously (Itkin et al., 2011, *supra*). In addition, the *TIP41* gene (*23*) was used as an endogenous control for the potato samples. Oligonucleotides are listed in **Table 3** hereinabove.

Production of recombinant enzyme

[00801] *GAME2, GAME17 and GAME18* were amplified from cDNA and subcloned into pACYCDUET-1 using *BamH* I and *Pst* I (*GAME2*, *GAME18*) or BamHI and XhoI (*GAME17*) restriction sites, and the insert was verified by sequencing. The resulting plasmids, pAC-*GAME2/17/18* were transformed to *E. coli* BL21 DE3. For expression of the GAME enzymes, fresh overnight cultures were diluted 1:100 in 25 ml 2xYT medium with 30 μg/ml chloramphenicol and incubated at 37°C and 250 rpm until an A_{600nm} of 0.4 was reached. Subsequently, IPTG was added to a concentration of 0.5 mM, and the incubation was continued overnight at 18°C and 250 rpm. The next day, cells were harvested by centrifugation, and the pellet resuspended in 2 ml of 50 mM Tris HCl pH = 7.0, 15% glycerol, 0.1 mM EDTA and 5 mM β-mercaptoethanol. After breaking the cells by sonication, insoluble material was removed by centrifugation, and the soluble fractions were used for characterization of the enzymes. Proteins were stored at -20°C until further analysis.

Preparation of substrates

[00802] For hydrolysis, 35 mg of α-tomatine was solved in 3 ml of 1N HCl, and was incubated for 15 min. at 100°C. Subsequently, the solution was put on ice, and NH_3 was added until the pH of the

solution was 9.0. The solution was extracted with 4 ml water-saturated butanol. The butanol phase was evaporated to dryness under vacuum, the residual pellet solved in 1 ml methanol and stored at -20°C until further use. The degradation products of α-tomatine were separated on a Luna 5 µm C18(2) 100 Å, LC Column 150 x 21.2 mm (Phenomenex, USA), using an isocratic elution with 25% acetonitrile in water and 0.1% formic acid. Compounds were detected using a 3100 Mass Detector (Waters), and collected. Fractions were freeze-dried, and purity of compounds was verified by LC-MS. For identification of products, liquid chromatography, coupled to quadrupole time-of-flight mass spectrometry (LC–QTOF-MS) was performed using a Waters Alliance 2795 HPLC connected to a Waters 2996 PDA detector and subsequently a QTOF Ultima V4.00.00 mass spectrometer (Waters, MS Technologies, UK) operated in positive ionization mode. The column used was an analytical Luna 3 µm C18 (2) 100 Å; 150 × 2.0 mm (Phenomenex, USA) attached to a C18 pre-column (2.0 × 4 mm; AJO-4286; Phenomenex, USA). Degassed eluent A [ultra-pure water:formic acid (1000:1, v/v)] and eluent B [acetonitrile:formic acid (1000:1, v/v)] were used with flow rate of 0.19 ml/min. The gradient started at 5% B and increased linearly to 75% B in 45 min., after which the column was washed and equilibrated for 15 min. before the next injection. The injection volume was 5 µl. This procedure yielded several milligrams of pure γ-tomatine (tomatidine - galactoside - glucoside, T-Gal-Glu) and $\beta 1$-tomatine (tomatidine - galactoside - diglucoside. T-Gal-Glu-Glu). Tomatidine galactoside (T-Gal) could not be purified in this way due to strong contamination with T-Gal-Glu. Therefore 5 mg tomatidine was incubated with GAME1 and UDP-galactose in 1 ml reaction mix, as described previously (Itkin et al., 2011, *supra*). T-Gal was purified from UDP-galactose by solid phase extraction. Waters OASIS HLB 3 cc columns (Waters Corp., Milford, MA) was conditioned with 6 mL 100% methanol followed by rinsing with 4 mL ultra-pure water. The reaction, supplemented with 10% methanol, was loaded and the cartridge was subsequently washed with 4 mL ultra-pure water. Compounds were eluted with 1 mL 75% methanol in ultra-pure water (v:v), and 0.4 mL 100% methanol. The solvent was removed from the combined eluate using a speed vacuum concentrator until a totally dry-pellet was obtained.

Enzyme assays

[00803] The substrates T-Gal, β1- and γ-tomatine were dissolved to 1 mM in 50% DMSO. Enzyme assays were carried out in 50 mM Tris HCl pH = 7.0 containing 5 mM β-mercaptoethanol using 5 µg/ml enzyme, 8 mM UDP-xylose and 0.02 mM substrate in a final reaction volume of 100 µl. After 2 h. of incubation under agitation at 37oC, reactions were stopped by addition of 300 µl methanol and 0.1% formic acid, and followed by brief vortexing and sonication for 15 min. Subsequently, the extracts were centrifuged for 5 min. at 13,000 rpm and filtered through 0.45 µm filters (Minisart SRP4, Biotech GmbH, Germany), and analyzed by LC-MS (see above). The amount of product was measured by the peak surface area in the LC-MS chromatogram, and compared to a control incubation

in which an enzyme preparation of an E. coli harboring an empty pACYCDUET-1. Masses used for detection were α-tomatine (C50H83NO21; m/z = 1034.55 ([M+H]+)), β1-tomatine T-Gal-Glu-Glu (C45H75NO17; m/z = 902.51 ([M+H])), β2-tomatine (C44H73NO16; m/z = 872.50 ([M+H]+)), γ-tomatine T-Gal-Glu (C39H65NO12; m/z = 740.46 ([M+H])), and T-Gal (C33H55NO7; m/z = 578.41 ([M+H])).

Virus Induced Gene Silencing (VIGS) experiments

[00804] Vectors containing fragments of *GAME* genes were constructed and VIGS experiments were conducted as described previously (Orzaea D et al., 2009. Plant Physiol. 150:1122–1134; Li R et al., 2006 J. Mass Spec. 41:1–22). Plants infected with Agrobacterium, containing empty vector and helper vector pTRV1, were used as control. Oligonucleotides used to prepare the pTRV2_DR_GW vectors are listed in **Table 3** hereinabove.

Genome sequence analysis of the wild tomato species

[00805] Partial genomic data obtained by re-sequencing (Dr. Arnaud G. Bovy, unpublished data) of three tomato wild species genomes (i.e. *Solanum pennellii*, *S. pimpinellifolium* and *S. chmielewskii*) were analyzed for the presence or absence of sequences (contigs) that align to the SGAs biosynthesis gene clusters on tomato chromosomes 7 and 12. The TopHat toolkit (Trapnell C. 2012. Nat. Protoc. 7:562–578) was used for mapping reads of the wild species to the tomato genome (ITAG 2.4), as a reference genome. The mapped reads were visualized with the IGV genome browser (Robinson J T et al., 2011. Nat. Biotechnol. 29:24–26). In order to assemble and align the sequence of the contigs from the three wild species to the gene clusters on to the existing cultivated tomato sequences of chromosomes 7 and 12, a combination of the CLC workbench, CAP3 BWA and SAMtools software packages and an in-house Perl script were used.

Example 2: Genes associated with SGA biosynthesis

[00806] To discover genes associated with SGA biosynthesis, a co-expression analysis using transcriptome data from tomato and potato plants was performed. Coexpression with GAME1/SGT1 (chromosome 7) and GAME4 (chromosome 12) as "baits" in either potato or tomato are presented in a form of a heatmap in **Tables 5-8** herein below. Genes that are highly co-expressed with either GAME1 /SGT1 (chromosome 7) or GAME4 (chromosome 12) are depicted with a large font and bold.

[00807] **Table 5: Accession numbers, putative protein and co-expression *r*-values – tomato, chromosome 7**

Gene name	Putative protein	*r*-value of correlation with tomato *GAME1* expression
Solyc07g043310	Aminotransferase	-0.26
Solyc07g043320	Unknown Protein	0.12
Solyc07g043330	GRAS family transcription factor	0.72
Solyc07g043340	Unknown Protein	
Solyc07g043350	Unknown Protein	
Solyc07g043360	60S ribosomal protein L27	0.10
Solyc07g043370	Transposase	
Solyc07g043380	Unknown Protein	
Solyc07g043390	Cellulose synthase family protein (GAME15)	0.92
Solyc07g043400	Unknown Protein	
Solyc07g043410	UDP-xylose xylosyltransferase (GAME2)	
Solyc07g043420	2-oxoglutarate-dependent dioxygenase	0.79
Solyc07g043430	Gag-Pol polyprotein	
Solyc07g043440	Glucosyltransferase-like protein	
Solyc07g043450	Zeatin O-glucosyltransferase	
Solyc07g043460	Cytochrome P450 (GAME 6)	0.91
Solyc07g043470	Unknown Protein	
Solyc07g043480	UDP-glucose glucosyltransferase	0.88
Solyc07g043490	UDP-glucosyltransferase family 1 protein (GAME1)	1.00
Solyc07g043500	UDP-glucosyltransferase	0.95
Solyc07g043510	Cysteine-type peptidase	-0.24
Solyc07g043520	transposase	
Solyc07g043530	Unknown Protein	
Solyc07g043540	Unknown Protein	
Solyc07g043550	UDP-arabinose 4-epimerase	0.70
Solyc07g043560	Heat shock protein 4	0.24
Solyc07g043570	Aldo/keto reductase family protein	-0.09
Solyc07g043580	BHLH transcription factor	0.43
Solyc07g043590	Amine oxidase family protein	0.03
Solyc07g043600	Pentatricopeptide repeat-containing protein	0.43
Solyc07g043610	Auxin response factor 6	

Gene name	Putative protein	*r*-value of correlation with tomato *GAME1* expression
Solyc07g043620	Auxin response factor 6-1	0.65
Solyc07g043630	Acyl-CoA synthetase/AMP-acid ligase II	
Solyc07g043640	Acyl-CoA synthetase/AMP-acid ligase II	
Solyc07g043650	AMP-dependent synthetase and ligase	
Solyc07g043660	Acyl-CoA synthetase/AMP-acid ligase II	-0.16
Solyc07g043670	Hydroxycinnamoyl CoA quinate transferase 2	
Solyc07g043680	Enoyl-CoA-hydratase	
Solyc07g043690	Enoyl-CoA-hydratase	
Solyc07g043700	Acyltransferase	

[00808] **Table 6: Accession numbers, putative protein and co-expression *r*-values – potato, chromosome 7**

Gene name	Putative protein	*r*-value of correlation with potato *SGT1* expression
PGSC0003DMG400011754	Gamma aminobutyrate transaminase	-0.31
PGSC0003DMG400011753	Uro-adherence factor A	-0.40
PGSC0003DMG400011742	DELLA protein RGA	0.15
PGSC0003DMG400011741	60S ribosomal protein L27	0.43
PGSC0003DMG400039612	Conserved gene of unknown function	
PGSC0003DMG400011752	Cellulose synthase (GAME15)	0.90
PGSC0003DMG400011740	beta-solanine rhamnosyltransferase (SGT3)	0.90
PGSC0003DMG400011751	2-oxoglutarate-dependent dioxygenase	0.87
PGSC0003DMG400011750	Cytochrome P-450 (GAME 6)	0.92
PGSC0003DMG400044993	Unknown Protein	
PGSC0003DMG400011749	solanidine galactosyltransferase (SGT1)	1.00
PGSC0003DMG402015928	OTU-like cysteine protease family protein	-0.24
PGSC0003DMG401015928	Conserved protein of unknown function	-0.25

Gene name	Putative protein	*r*-value of correlation with potato *SGT1* expression
PGSC0003DMG400015927	UDP-arabinose 4-epimerase 1	-0.21
PGSC0003DMG400015920	Heat shock 70 kDa protein	-0.17
PGSC0003DMG402015926	Aldo/keto reductase	-0.05
PGSC0003DMG401015926	Isoform 2 of Transcription factor PIF5	-0.33
PGSC0003DMG400015925	Amine oxidase	0.11
PGSC0003DMG400015924	Pentatricopeptide repeat-containing protein	0.32
PGSC0003DMG400015919	ARF8	0.07
PGSC0003DMG400036440	AMP dependent ligase	
PGSC0003DMG400015923	Acyl:coA ligase acetate-coA synthetase	
PGSC0003DMG400015922	Acyl:coA ligase acetate-coA synthetase	
PGSC0003DMG400044288	Acyltransferase	
PGSC0003DMG400015918	Acyltransferase	0.03

[00809] **Table 7: Accession numbers, putative protein and co-expression *r*-values – tomato, chromosome 12**

Gene name	Putative protein	*r*-value of correlation with tomato *GAME4* expression
Solyc12g006530	Cycloartenol synthase	0.08
Solyc12g006520	Cycloartenol synthase	0.05
Solyc12g006510	Cycloartenol Synthase	-0.12
Solyc12g006500	Phosphate translocator protein	0.15
Solyc12g006490	Beta-1-3-galactosyl-o-glycosyl-glycoprotein	0.03
Solyc12g006480	Nup205 protein	0.35
Solyc12g006470	gamma-aminobutyrate Aminotransferase-like protein	0.94
Solyc12g006460	Cytochrome P450 (GAME 4)	1.00
Solyc12g006450	gamma-aminobutyrate Aminotransferase-like protein	-0.13
Solyc12g006440	Unknown Protein	0.25
Solyc12g006430	UDP-glucuronosyltransferase 1-1 82A1	
Solyc12g006420	Topoisomerase II-associated	0.08

Gene name	Putative protein	*r*-value of correlation with tomato *GAME4* expression
	protein PAT1	
Solyc12g006410	UDP-arabinse 4-epimerase	
Solyc12g006400	Unknown Protein	
Solyc12g006390	2-oxoglutarate-dependent dioxygenase	
Solyc12g006380	2-oxoglutarate-dependent dioxygenase	0.15
Solyc12g006370	Amine oxidase family protein	-0.16
Solyc12g006360	Multidrug resistance protein mdtK	
Solyc12g006350	Auxin response factor 6	0.35
Solyc12g006340	Auxin response factor 6	0.47
Solyc12g006330	Acyltransferase-like protein	
Solyc12g006320	ATP-dependent RNA helicase	0.14
Solyc12g006310	Endoplasmic reticulum-Golgi	0.25
Solyc12g006300	WD-repeat protein-like	-0.03
Solyc12g006290	Reticulon family protein	0.19
Solyc12g006280	Myb-like DNA-binding protein	

[00810] **Table 8: Accession numbers, putative protein and co-expression *r*-values – potato, chromosome 12**

Gene name	Putative protein	*r*-value of correlation with potato *GAME4* expression
PGSC0003DMG400020034	Beta-amyrin synthase	-0.13
PGSC0003DMG400024276	Beta-Amyrin Synthase	-0.09
PGSC0003DMG400024277	Gene of unknown function	0.10
PGSC0003DMG400024278	Phenylacetaldehyde synthase	0.10
PGSC0003DMG400024279	Conserved gene of unknown function	-0.16
PGSC0003DMG400024280	Triose phosphate/phosphate translocator, non-green plastid, chloroplast	-0.06
PGSC0003DMG400024271	Acetylglucosaminyltransferase	-0.06
PGSC0003DMG400024273	Resistance protein PSH-RGH6	0.37
PGSC0003DMG400024281	Gamma aminobutyrate transaminase isoform2	0.94
PGSC0003DMG400024274	Cytochrome P450	1.00

Gene name	Putative protein	*r*-value of correlation with potato *GAME4* expression
	monooxygenase GAME4	
PGSC0003DMG400024275	Gamma aminobutyrate transaminase isoform3	0.37
PGSC0003DMG400024282	Fortune-1	0.36
PGSC0003DMG400028806	UDP-glycosyltransferase 82A1-like	-0.18
PGSC0003DMG401028807	Topoisomerase II-associated protein PAT1	
PGSC0003DMG402028807	UDP-arabinse 4-epimerase	
PGSC0003DMG400028824	Gene of unknown function	
PGSC0003DMG400028808	2-oxoglutarate-dependent dioxygenase	-0.07
PGSC0003DMG400028809	2-oxoglutarate-dependent dioxygenase	0.61
PGSC0003DMG400028810	Amine oxidase	-0.04
PGSC0003DMG400028825	MATE transporter	
PGSC0003DMG400028826	Auxin response factor 6	
PGSC0003DMG400043090	Integrase core domain containing protein	
PGSC0003DMG400037700	WRKY transcription factor 27	
PGSC0003DMG400028811	Acyltransferase	
PGSC0003DMG400028812	DEAD-box ATP-dependent RNA helicase 53	0.56
PGSC0003DMG400028814	WD-repeat protein	-0.10
PGSC0003DMG401028829	Polygalacturonase	
PGSC0003DMG400028815	Reticulon family protein	0.08
PGSC0003DMG400028830	Myb-like DNA-binding domain, SHAQKYF class family protein	

[00811] Sixteen genes from each species were co-expressed with *GAME1/SGT1* (**Table 9**, **FIGURE 2**). One of these genes, previously designated *GLYCOALKALOID METABOLISM 4* (*GAME4*), encodes a member of the 88D subfamily of cytochrome P450 proteins (**FIGURE 3**). *GAME4* and *GAME1/SGT1* display a very similar expression profile in tomato and potato (WO 2010/095843). The *GAME1/SGT1* and *GAME4* genes in tomato and potato are positioned in chromosomes 7 and 12 such that they are physically next to several of their co-expressed genes (**FIGURE 2**).

[00812] A cluster of *GAME1/SGT1* co-expressed genes spans a ~200Kbp genomic region on

chromosome seven. Together with *GAME1,* the tomato cluster is composed of 7 co-expressed genes. These include 3 UDP-glycosyltransferases [*GAME2* (termed *SGT3 in potato*); *GAME17* and *GAME18*], a cytochrome P450 of the 72A subfamily (*GAME6*), a 2-oxoglutarate-dependent dioxygenase (*GAME11*), and a cellulose synthase-like protein (GAME15). It appears that in potato this cluster contains 5 co-expressed genes as it lacks homologs of the tomato genes encoding GAME17 and GAME18 UDP-glycosyltransferases. Enzyme activity assays were performed with the four recombinant clustered tomato UDP-glycosyltransferases. GAME17 and GAME18 exhibited UDP-glucosyltransferase activity when incubated with tomatidine galactoside (T-Gal) and γ-tomatine (T-Gal-Glu) as a substrate, respectively, whereas GAME2 was shown to have an UDP-xylosyltransferase activity when incubated with $\beta 1$-tomatine (T-Gal-Glu-Glu) as a substrate (**FIGURES 4E, 4F, and 4G**). GAME1 was previously shown to act as a tomatidine UDP-galactosyltransferase in tomato (Itkin et al., 2011, *supra).* When incubating the 4 recombinant UGT enzymes in a single test tube, with tomatidine, and all glycoside donors (UDP -galactose, -glucose and -xylose), the accumulation of the final SGA product α-tomatine was observed (**FIGURE 4H**).

[00813] Two genes encoding putative transcription factors were identified among the genes co-expressed with *GAME1/SGT1* and *GAME4* (**FIGURE 4**): one gene, designated *GAME9*, was identified by the tomato ID Solyc01g090340 and by the potato ID PGSC0003DMG400025989. It is described as ethylene-responsive element binding factor 13 and contains a putative AP2 domain. The other gene is the BHLH-transcription factor, identified by the tomato ID Solyc03g046570 and by the potato ID PGSC0003DMG400012262.

[00814] **Table 9: Details of homologs co-expressed with known and putative steroidal alkaloid-associated genes in both potato and tomato presented in FIGURE 2**

[00815] Tomato and potato sequences were obtained from Sol Genomics Network (solgenomics.net). r-value for co-expression ≥ 0.8. TCON number, a contig reference name given by the inventors in the assembly of RNAsec data. N/A, not available.

Name	Tomato ID Solyc	Potato reads	Tomato ID
Extensin-like protein	Solyc01g006400	PGSC0003DMG400023230	TCONS_00007692
GAME 9	Solyc01g090340 [amino acid SEQ ID NO: 13] [nucleic acid SEQ ID NO: 15]	PGSC0003DMG400025989 [amino acid SEQ ID NO: 14] [nucleic acid SEQ ID NO: 16]	TCONS_00011729
Delta (24)-sterol reductase-like	Solyc02g069490	PGSC0003DMG400021142	TCONS_00044548
BHLH transcription factor	Solyc03g046570 [amino acid SEQ ID NO:17] [nucleic acid SEQ ID	PGSC0003DMG400012262 [nucleic acid SEQ ID NO: 20] [nucleic acid SEQ ID NO:	TCONS_00055879

	NO: 18] [nucleic acid SEQ ID NO: 19]	21]	
LRR receptor-like protein kinase	Solyc05g009100	PGSC0003DMG400014576	TCONS_00101281
Glycosyltransferase	Solyc05g053120	PGSC0003DMG402027210	TCONS_00100675
Cellulose synthase-like (GAME15)	Solyc07g043390	PGSC0003DMG400011752	TCONS_00135034
GAME6 (CYP72)	Solyc07g043460	PGSC0003DMG400011750	TCONS_00137734
GAME1 (Galactosyltransferase)	Solyc07g043490	PGSC0003DMG400011749	TCONS_00133014
GAME7 (CYP72)	Solyc07g062520 (*GAME1* *r*-value 0.66; *GAME4* *r*-value 0.71)	PGSC0003DMG402012386 (*SGT1* *r*-value 0.63; *GAME4* *r*-value 0.73)	TCONS_00132326
Srt/Thr protein kinase 6	Solyc08g066050	PGSC0003DMG400025461	TCONS_00151251
Meiotic serine proteinase	Solyc08g077860	PGSC0003DMG401012339	TCONS_00149157
Sterol reductase	Solyc09g009040	PGSC0003DMG400002720	TCONS_00162820
Ubiquitin protein ligase	Solyc10g008410	PGSC0003DMG400021683	TCONS_00183263
Proteinase inhibitor II	Solyc11g020960	PGSC0003DMG402003479	TCONS_00194999
GAME4 (CYP88)	Solyc12g006460	PGSC0003DMG400024274	TCONS_00210154
Gamma-aminobutyrate Aminotransferase-like protein (transaminase) (GAME12)	Solyc12g006470	PGSC0003DMG400024281	
Beta-solanine rhamnosyltransferase (SGT3)	#N/A	PGSC0003DMG400011740	
2-oxoglutarate-dependent dioxygenase (GAME11)	Solyc07g043420 [amino acid SEQ ID NO:22] [nucleic acid SEQ ID NO: 24] [nucleic acid SEQ ID NO: 25]	PGSC0003DMG400011751 [amino acid SEQ ID NO:23] [nucleic acid SEQ ID NO: 26] [nucleic acid SEQ ID NO: 27]	
GAME18 (Glycosyltransferase)	Solyc07g043500	#N/A	
GAME17 (Glycosyltransferase)	Solyc07g043480	#N/A	

Example 3: Functional analysis of GAME9-transcription factor

[00816] GAME9-silencing (RNAi) and overexpressing (OX) constructs were created by introducing the corresponding *GAME9* DNA fragments to pK7GWIWG2(II) and pJCV52 binary vectors, respectively. Transgenic tomato and potato lines transformed with the respective GAME9 silencing and overexpressing constructs were generated as previously described (Itkin et al., 2011, *supra*). Tissue extracts were prepared and analyzed as described in Itkin et al. (2011, *supra*). SEQ ID NO: 18 presents the sequence of the GAME9 RNAi silencing molecule. The metabolic profiling of steroidal alkaloids using UPLC-TQ-MS was performed on extracts obtained from leaves and/or

tubers of transgenic and wild type tomato and/or potato plants. In extract obtained from potato tuber peels of potato lines in which the gene encoding GAME9 was silenced (GAME9-RNAi lines) a reduction in α-solanine and α-chaconine was observed (**FIGURES 5A and 5B**, respectively). Leaves from potato GAME9-overexpression lines contained higher levels of α-solanine (**FIGURE 5C**) and α-chaconine (**FIGURE 5D**) compared to the wild type. A similar accumulation pattern was observed in potato leaves, having reduced amounts of α-chaconine and α-solanine in RNAi lines and increased amounts of these steroidal alkaloids in lines overexpressing the GAME9-transcription factor (**FIGURE 6**).

[00817] In tomato, leaves extract of a line overexpressing the GAME9-transcription factor (designated 5879) contained higher levels of α-tomatine compared to its amount in leaf extract obtained from wild type plants. On the contrary, down regulation of the expression of GAME9-transcription factor (line 5871) resulted in significant reduction of α-tomatine content.

Example 4: Functional characterization of the GAME genes

[00818] *GAME11 silenced plants*

[00819] Virus induced gene silencing (VIGS) is a commonly used technique allowing systemic silencing of genes in various organs of the plant (Dinesh-Kumar S P et al., 2003. Methods Mol Biol 236:287-294). SEQ ID NO: 19 presents the sequence of the GAME11 RNAi silencing molecule.

[00820] Analysis of tomato leaves with VIGS-silenced *GAME11*, a putative dioxygenase in the cluster, revealed a significant reduction in α-tomatine levels and accumulation of several cholestanol-type steroidal saponins.

[00821] Silencing of *GAME11* dioxygenase in tomato results in depletion of α-tomatine levels in leaves (m/z = 1034.5) (**FIGURE 8A**) while accumulating cholestanol-type steroidal saponins [i.e. STSs; m/z = 1331.6, 1333.6, 1199.6, 1201.6 (major saponins)] (**FIGURE 8B**). **FIGURE 8C** shows MS/MS spectrum of m/z = 1331.6 (at 19.28 min.). **FIGURE 8D** shows the fragmentation patterns of the saponin eluted at 19.28 min. and accumulating in *GAME11*-silenced leaves. The corresponding mass signals are marked with an asterisk on the MS/MS chromatogram in **FIGURE 8C**. The elemental composition and fragmentation patterns show that the compounds are cholestanol-type saponins, lacking one hydroxy-group and the E-ring (in comparison to furostanol-type saponins), which results in fragmentation, involving multiple losses of water molecules instead of tautomerisation and McLafferty rearrangement of the *E*-ring.

[00822] *GAME18* silenced plants

[00823] The role of *GAME18* in creating the tetrasaccharide moiety of α-tomatine was supported by Virus Induced Gene Silencing (VIGS) assays as *GAME18*-silenced fruit accumulated γ-tomatine which was not present in the control sample (**FIGURE 9**).

[00824] Among the metabolites extracted from *GAME18*-silenced mature green fruit, peaks of newly accumulating compounds were detected, corresponding to the γ-tomatine standard (m/z = 740.5) (**FIGURES 9A-C**), and γ-tomatine pentoside (m/z = 872.5) (**FIGURES 9D-9E**).

[00825] *GAME12* silenced plants

[00826] Silencing of *GAME12* transaminase in tomato resulted in accumulation of a furastanol-type steroidal saponin (**FIGURE 4D**). **FIGURE 10A** shows that *GAME12*-silenced leaves accumulate an STS (m/z=753.4), while it exists in only minor quantities in wild type leaf **FIGURE 10B**. **FIGURE 10C** shows MS/MS spectrum of m/z=753.4 at 19.71 min. with interpretation of the fragments. Suggested structure of the STS at 19.71 min. is depicted in **FIGURE 10D**, concluded from the characteristic mass fragments observed in the MS/MS experiment.

[00827] *Function of GAME7 and GAME8*

[00828] Genes that were tightly co-expressed and positioned elsewhere in the genome were also functionally examined. Two genes, designated *GAME7* and *GAME8* belong to the CYP72 subfamily of cytochrome P450s. *GAME7* was co-expressed in both species (potato and tomato) while StGAME8a and StGAME8b were strongly co-expressed with *StSGT1* and *StGAME4* in potato. At present, we could not demonstrate SGA-related activity for GAME7 although as for GAME6 it was suggested to be involved in SGA metabolism (US 20120159676). Yet, *GAME8*-silenced tomato leaves accumulated 22-(*R*)-hydroxycholesterol (**FIGURES 11A-11D**), a proposed intermediate in the SGA biosynthetic pathway (**FIGURE 1**). *GAME8*-silenced line accumulates both isomers in comparison to wild type (**FIGURE 11D**). The (R)-isomer is more abundant and hence most likely to be the substrate of GAME8.

[00829] **FIGURE 12** shows the phylogenetic tree of GAME genes in the plant CYP450 protein family. The numbers on the branches indicate the fraction of bootstrap iterations supporting each node.

Example 5: Proposed biosynthetic pathway in Solanaceous plants

[00830] An expanded biosynthetic pathway in *Solanaceous* plants has been proposed, as depicted in the schematic of **FIGURE 13** (dashed arrows represent multiple enzymatic reactions in the pathway) with respect to the tomato. This pathway can be broken down into four parts for convenience. In Part I, a series of reactions (catalyzed, e.g., by SSR2, SMO3, SM04) converts cylcloartenol to cholesterol. Byproducts include triterpenoids and phytosterols. In Part II, a series of reactions (catalyzed, e.g., by GAME11, GAME6, GAME4, GAME12, GAME25) converts cholesterol to tomatidine (aglycone). Byproducts include steroidal saponins (e.g., uttroside B). In Part III, a series of reactions (catalyzed, e.g., by GAME1, GAME 17, GAME18, GAME2) converts

tomatidine to steroidal glycoalkaloids (e.g., α-tomatine). In Part IV, a series of reactions converts steroidal glycoalkaloids (e.g., α-tomatine) of a green tomato to lycoperosides and/or esculeosides (e.g., esculeoside A) of a red tomato.

Example 6. Pathways involving steroidal glycoalkaloid (SGA) biosynthesis in tomato, potato, and eggplant

[00831] A cellulose synthase-like gene (GAME15) in tomato, potato, and eggplant has been identified as being associated with steroidal glycoalkaloid (SGA) biosynthesis (**FIGURES 14A-14C**). This gene has been shown to have been strongly co-expressed with other SGA biosynthesis genes (e.g., GAME4, GAME12) and also with regulators of SGA biosynthesis (e.g., GAME9). **FIGURE 14D** provides a non-limiting example of a proposed pathway in *Solanaceous* plants, for example but not limited to, wherein a skilled artisan would recognize that the pathway encompasses production of tomatidine 3-O-glucoronide from tomatidine by a GAME15 glucuronic acid transferase activity. In certain embodiments, the GAME15 enzyme comprises a GAME15 present in a Solanaceous plant, for example but not limited to *Solanum tuberosum*, potato; *Solanum lycopersicum*, tomato; *Solanum dulcamara*, bitter-sweet; and *Solanum melogena*, eggplant.

[00832] Sequences were identified as follows:

[00833] Cellulose synthase like_tomato

[00834] ATGAAAAAAACCATGGAGCTCAACAAAAGCACTGTTCCACAACCTATCACC
ACCGTATACCGACTCCACATGTTCATCCACTCAATAATCATGCTTGCATTAATATACTA
CCGTGTATCTAATTTGTTTAAATTCGAAAACATTCTCAGTTTACAAGCACTTGCTTGGG
CGCTCATCACTTTTGGTGAATTTAGTTTCATTCTCAAGTGGTTCTTCGGACAAGGTACT
CGTTGGCGCCCCGTTGAACGAGATGTTTTCCCTGAAAACATTACTTGCAAAGATTCCG
ATCTACCGCCAATTGACGTAATGGTATTCACTGCCAATCCTAAGAAAGAGCCAATTGT
AGATGTCATGAACACTGTGATATCCGCAATGGCTCTTGATTATCCCACCGATAAATTG
GCTGTGTATCTCGCTGATGATGGAGGATGTCCATTGTCGTTGTACGCCATGGAACAAG
CGTGTTTGTTTGCAAAGCTATGGTTACCTTTCTGTAGAAACTATGGAATTAAAACGAG
ATGCCCAAAAGCATTTTTTTCTCCGTTAGGAGATGATGACCGTGTTCTTAAGAATGAT
GATTTTGCTGCTGAAATGAAAGAAATTAAATTGAAATATGAAGAGTTCCAGCAGAAG
GTGGAACATGCTGGTGAATCTGGAAAAATCAATGGTAACGTAGTGCCTGATAGAGCT
TCGCTTATTAAGGTAATAAACGAGAGGGAGAACGAAAAGAGTGTGGATGATATGACG
AAAATGCCCTTGCTAGTTTATGTATCCCGTGAAAGAAGATTCAACCGTCTTCATCATTT
CAAGGGTGGATCTGCAAATGCTCTACTTCGAGTTTCTGGAATAATGAGTAATGCCCCC
TATGTACTGGTGTTAGATTGTGATTTCTTCTGTCATGATCCAATATCAGCTAGGAAGGC

AATGTGTTTTCATCTTGATCCAAAGCTATCATCTGATTTAGCCTATGTTCAGTTCCCTC
AAGTCTTTTACAATGTCAGCAAGTCAGATATTTATGATGTCAAAATTAGACAGGCTTA
CAAGACAATATGGCATGGAATGGATGGTATCCAAGGCCCAGTGTTATCTGGGACTGG
TTATTTTCTCAAGAGGAAAGCGTTATACACAAGTCCAGGAGTAAAAGAGGCGTATCTT
AGTTCACCGGAAAAGCATTTTGGAAGGAGTAAAAGGTTTCTTGCTTCATTAGAGGAG
AAAAATGGTTATGTTAAGGCAGATAAAGTCATATCAGAAGATATCATAGAGGAAGCT
AAGATGTTAGCTACTTGTGCATATGAGGATGGCACACATTGGGGTCAAGAGATTGGTT
ATTCATACGATTGTCATTTGGAGAGCACTTTTACTGGTTATCTATTACACTGCAAAGG
GTGGACATCTACTTATTTGTATCCAGACAGGCCATCTTTCTTGGGTTGTGCCCCAGTTG
ATATGCAAGGTTTCTCATCACAGCTCATCAAATGGGTTGCTGCACTTACACAAGCTGG
TTTATCACATCTCAATCCCATCACTTATGGTTTGAGTAGTAGGATGAGGACTCTCCAAT
GCATGTGCTATGCCTATTTGATGTATTTCACTCTTTATTCTTGGGGAATGGTTATGTAT
GCTAGTGTTCCTTCTATTGGCCTTTTGTTTGACTTCCAAGTCTATCCTGAGGTACATGA
TCCGTGGTTTGCAGTGTATGTGATTGCTTTCATATCGACAATTTTGGAGAATATGTCGG
AGTCAATTCCAGAAGGGGGATCAGTTAAAACGTGGTGGATGGAATACAGGGCATTGA
TGATGATGGGAGTTAGCGCAATATGGTTAGGAGGATTGAAAGCTATATATGACAAGA
TAGTCGGAACACAAGGAGAGAAATTGTATTTGTCGGACAAGGCAATTGACAAGGAAA
AGCTCAAGAAATACGAGAAGGGCAAATTTGATTTCCAAGGAATAGGGATACTTGCTC
TGCCACTGATAGCATTTTCCGTGTTGAACCTCGTAGGCTTCATTGTTGGAGCTAATCAT
GTCTTTATTACTATGAACTACGCAGGCGTGCTGGGCCAACTCCTCGTATCATCGTTCTT
CGTCTTTGTTGTCGTCACTGTTGTCATTGATGTTGTATCTTTCTTAAAGGTTTCTTAA
[SEQ ID NO: 30]

[00835] Cellulose synthase like (tomato)

[00836] MKKTMELNKSTVPQPITTVYRLHMFIHSIIMLALIYYRVSNLFKFENILSLQALA
WALITFGEFSFILKWFFGQGTRWRPVERDVFPENITCKDSDLPPIDVMVFTANPKKEPIVDV
MNTVISAMALDYPTDKLAVYLADDGGCPLSLYAMEQACLFAKLWLPFCRNYGIKTRCPK
AFFSPLGDDDRVLKNDDFAAEMKEIKLKYEEFQQKVEHAGESGKINGNVVPDRASLIKVI
NERENEKSVDDMTKMPLLVYVSRERRFNRLHHFKGGSANALLRVSGIMSNAPYVLVLDC
DFFCHDPISARKAMCFHLDPKLSSDLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWHGMD
GIQGPVLSGTGYFLKRKALYTSPGVKEAYLSSPEKHFGRSKRFLASLEEKNGYVKADKVIS
EDIIEEAKMLATCAYEDGTHWGQEIGYSYDCHLESTFTGYLLHCKGWTSTYLYPDRPSFL
GCAPVDMQGFSSQLIKWVAALTQAGLSHLNPITYGLSSRMRTLQCMCYAYLMYFTLYSW
GMVMYASVPSIGLLFDFQVYPEVHDPWFAVYVIAFISTILENMSESIPEGGSVKTWWMEY
RALMMMGVSAIWLGGLKAIYDKIVGTQGEKLYLSDKAIDKEKLKKYEKGKFDFQGIGIL

ALPLIAFSVLNLVGFIVGANHVFITMNYAGVLGQLLVSSFFVFVVVTVVIDVVSFLKVS

[SEQ ID NO: 31]

[00837] Cellulose synthase like (*solanum pennellii*)

[00838] ATGAAAAAAACCATGGAGCTCAACAAAAGCACTGTTCCACAACCTATCACC
ACCGTATACCGACTCCACATGTTCATCCACTCAATAATCATGCTTGCATTAATATACTA
CCGTGTATCTAATTTGTTTAAATTCGAAAACATTCTCAGTTTACAAGCACTTGCTTGGC
TACTCATCACTTTTGGTGAATTTAGTTTCATTCTCAAGTGGTTCTTCGGACAAGGAACT
CGTTGGCGCCCCGTTGAACGAGATGTTTTCCCTGAAAACATTACTTGCAAAGATTCCG
ATCTACCGCCAATTGACGTAATGGTGTTCACTGCCAATCCTAAGAAAGAGCCAATTGT
AGATGTCATGAACACTGTGATATCCGCAATGGCTCTTGATTATCCCACCGATAAATTG
GCTGTGTATCTGGCCGATGATGGAGGATGTCCATTGTCCTTGTACGCCATGGAACAAG
CATGTTTGTTTGCAAAGCTATGGTTACCTTTCTGTAGAAAGTATGGAATTAAAACGAG
ATGCCCAAAAGCATTTTTTTCTCCGTTAGGAGATGATGACCGTGTTCTTAAGAATGAT
GATTTTGCTGCTGAAATGAAAGAAATTAAATTGAAATATGAAGAGTTCCAGCAGAAC
GTGGAACATGCTGGTGAATCTGGAAAAATCAATGGCAACGTAGTGCCTGACAGAGCT
TCGCTTATTAAGGTAATAAACGAGAGGGAGAACGAAAAGAGTGTCGATGATTTAACG
AAAATGCCCTTGCTAGTTTATGTATCCCGTGAAAGAAGATTCAACCGTCTTCATCATTT
CAAGGGTGGATCTGCAAATGCTCTACTTCGAGTTTCTGGAATAATGAGTAATGCCCCC
TATGTACTGGTGTTAGATTGTGATTTCTTCTGTCATGATCCGATATCAGCTAGGAAAGC
AATGTGTTTTCATCTTGATCCAAAGCTATCATCTGATTTAGCCTATGTTCAGTTCCCTC
AAGTCTTTTACAATGTCAGCAAGTCCGATATTTATGATGTCAAAATTAGACAGGCTTA
CAAGACAATATGGCATGGAATGGATGGTATCCAAGGCCCAGTGTTATCTGGAACTGG
TTATTTTCTCAAGAGGAAGGCGTTATACACAAGTCCAGGAGTAAAAGAGGCGTATCTT
AGTTCACCGGAAAAGCATTTTGGAAGGAGTAAAAAGTTCCTTGCTTCATTAGAGGAG
AAAAATGGTTATGTTAAGGCAGATAAAGTCATATCAGAAGATATCATAGAGGAAGCT
AAGATCTTAGCTACTTGTGCATATGAGGATGGCACACATTGGGGTCAAGAGATTGGTT
ATTCATACGATTGTCATTTGGAGAGCACTTTTACTGGTTATCTATTACACTGCAAAGG
GTGGACATCTACTTATTTGTATCCAGACAGGCCATCTTTCTTGGGTTGTGCCCCAGTTG
ATATGCAAGGTTTCTCATCACAGCTCATAAAATGGGTTGCTGCACTTACACAAGCTGG
TCTATCACATCTCAATCCCATCACTTATGGTTTGAGTAGTAGGATGAGAACTCTCCAA
TGCATGTGCTATGCCTATTTGATGTATTTCACTCTTTATTCTTGGGGAATGGTTATGTA
TGCTAGTGTTCCTTCTATTGGCCTTTTGTTTGGCTTCCAAGTCTACCCTGAGGTACATG
ATCCATGGTTTGCAGTGTATGTGATTGCTTTCATATCGACAATTTTGGAGAATATGTCG
GAGTCAATTCCAGAAGGGGGATCAGTTAAAACGTGGTGGATGGAATACAGGGCATTG

ATGATGATGGGAGTTAGCGCAATATGGTTAGGAGGATTGAAAGCTATATATGACAAG
ATAGTCGGAACACAAGGAGAGAAATTGTATTTGTCGGACAAGGCAATTGACAAGGAA
AAGCTCAAGAAATACGAGAAGGGCAAATTTGATTTCCAAGGAATAGGGATACTTGCT
CTGCCATTGATAGCATTTTCCGTGTTGAACCTCGTAGGCTTCATTGTTGGAGCTAATCA
TGTCTTTATTACTATGAACTACGCAGGCGTGCTGGGCCAACTCCTCGTATCATCATTCT
TCGTCTTTGTTGTCGTCACTGTTGTCATTGATGTTGTATCTTTCTTAAAGGTTTCTTAA
[SEQ ID NO: 32]

[00839] Cellulose synthase like (*solanum pennellii*)

[00840] MKKTMELNKSTVPQPITTVYRLHMFIHSIIMLALIYYRVSNLFKFENILSLQALA
WLLITFGEFSFILKWFFGQGTRWRPVERDVFPENITCKDSDLPPIDVMVFTANPKKEPIVDV
MNTVISAMALDYPTDKLAVYLADDGGCPLSLYAMEQACLFAKLWLPFCRKYGIKTRCPK
AFFSPLGDDDRVLKNDDFAAEMKEIKLKYEEFQQNVEHAGESGKINGNVVPDRASLIKVI
NERENEKSVDDLTKMPLLVYVSRERRFNRLHHFKGGSANALLRVSGIMSNAPYVLVLDC
DFFCHDPISARKAMCFHLDPKLSSDLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWHGMD
GIQGPVLSGTGYFLKRKALYTSPGVKEAYLSSPEKHFGRSKKFLASLEEKNGYVKADKVIS
EDIIEEAKILATCAYEDGTHWGQEIGYSYDCHLESTFTGYLLHCKGWTSTYLYPDRPSFLG
CAPVDMQGFSSQLIKWVAALTQAGLSHLNPITYGLSSRMRTLQCMCYAYLMYFTLYSWG
MVMYASVPSIGLLFGFQVYPEVHDPWFAVYVIAFISTILENMSESIPEGGSVKTWWMEYR
ALMMMGVSAIWLGGLKAIYDKIVGTQGEKLYLSDKAIDKEKLKKYEKGKFDFQGIGILAL
PLIAFSVLNLVGFIVGANHVFITMNYAGVLGQLLVSSFFVFVVVTVVIDVVSFLKVS [SEQ
ID NO: 33]

[00841] Cellulose synthase like (potato)

[00842] ATGAAAAAAACCATGGAGCTCAACAAAAGCACTGTTCCACAACCTATCACC
ACCATATACCGACTCCACATGTTTATCCACTCTATAATCATGGTTGCATTAATATACTA
CCGTGTATCTAATTTGTTTAAATTCGAAAACATTCTGAGTTTACAAGCACTTGCTTGGG
TACTCATCACTTTTGGTGAATTTAGTTTCATTCTCAAGTGGTTCTTCGGACAAGGAACT
CGTTATCGCCCTGTTGAAAGAGATGTTTTCCCTGAAAACATAACTTGCAAAGATTCCG
ATCTACCACCAATTGACGTAATGGTATTCACTGCCAATCCTAAGAAAGAGCCAATTGT
GGATGTCATGAACACTGTGATATCCGCAATGGCTCTTGATTATCCTACGGATAAATTG
GCTGTGTATCTGGCTGATGATGGAGGATGTCCTTTGTCATTGTACGCCATGGAAGAAG
CATGTGTGTTTGCAAAGCTGTGGCTACCTTTCTGTAGGAAGTATGGAATTAAAACTAG
ATGCCCTAAAGCGTTTTTTTCTCCTTTAGGAGATGATGAACGTGTTCTTAAGAATGATG
ATTTTGATGCTGAAATGAAAGAAATTAAATTGAAATATGAAGAGTTCCAGCAGAATG
TGGAACGTGCTGGTGAATCTGGAAAAATCAATGGTAACGTAGTGCCTGATAGAGCCT

CGTTTATTAAGGTAATAAACGACAGAAAAGCGGAGAGCGAAAAGAGTGCCGATGATT
TAACGAAAATGCCCTTGCTAGTTTATGTATCCCGTGAAAGAAGATTCAACCGTCTTCA
TCACTTCAAGGGTGGATCTGCAAATGCTCTTCTTCGAGTTTCTGGAATAATGAGTAAT
GCCCCCTATATACTGGTGTTAGATTGTGATTTCTTCTGTCATGATCCAATATCAGCTAG
GAAGGCAATGTGTTTTCATCTTGATCCAAAGCTATCATCTGATTTAGCTTATGTTCAGT
TCCCTCAAGTCTTTTACAATGTCAGCAAGTCCGATATTTATGATGTCAAAATTAGACA
GGCTTACAAGACAATATGGCATGGAATGGATGGTATCCAAGGCCCAGTGTTATCAGG
AACTGGTTATTTTCTGAAGAGGAAGGCGTTATACACGAGTCCAGGAGTAAAGGAGGA
GTATCTTAGTTCACCGGAAAAGCATTTTGGAAGGAGTAAAAAGTTCCTTGCTTCACTA
GAGGAGAAAAATGGTTATGTTAAGGCAGAGAAAGTCATATCAGAAGATATCGTAGAG
GAAGCTAAGACCTTAGCTACTTGTGCATATGAGGATGGCACACATTGGGGTCAAGAG
ATTGGTTATTCATACGATTGTCATTTGGAGAGCACTTTTACTGGTTATCTATTACACTG
CAAAGGGTGGAGATCGACTTATTTGTATCCAGACAGGCCATCTTTCTTGGGTTGTGCC
CCAGTTGATATGCAAGGTTTCTCCTCACAGCTCATAAAATGGGTTGCTGCACTTACAC
AAGCTGGTTTATCACATCTCAATCCCATCACTTATGGCTTTAGTAGCAGGATGAAAAC
TCTCCAATGCATGTGCTATGCCTATTTGATATATTTCACTCTTTATTCTTGGGGAATGG
TTCTATATGCTAGTGTTCCTTCTATTGGCCTTTTGTTTGGCTTCCAAGTCTATCCCGATG
TACATGATCCATGGTTTGCAGTGTATGTGATTGCTTTCATATCGGCAATTTTGGAGAAT
ATGTCGGAGTCAATTCCTGATGGGGGATCATTTAAATCTTGGTGGATGGAATACAGGG
CACTGATGATGATGGGAGTTAGTGCAATATGGTTAGGAGGATTGAAAGCTATATTAG
ACAGGATAATCGGAACAGAAGGAGAGAAATTGTATTTATCGGACAAGGCAATTGACA
AGGAAAAGCTCAAGAAATACGAGAAGGGGAAATTTGATTTCCAAGGAATAGGGATA
CTTGCTGTACCATTGATAGCATTTTCCTTGTTGAACCTCGTAGGCTTCATTGTTGGAGC
TAATCATGTCTTTATTACTATGAACTACGCAGGTGTGCTTGGCCAACTCCTCGTATCAT
CCTTCTTCGTCTTTGTCGTGGTCACTGTTGTCATTGATGTCGTTTCTTTCTTAAAGGTTT
CTTAA [SEQ ID NO: 34]

[00843] Cellulose synthase like (potato)

[00844] MELNKSTVPQPITTIYRLHMFIHSIIMVALIYYRVSNLFKFENILSLQALAWVLITF
GEFSFILKWFFGQGTRYRPVERDVFPENITCKDSDLPPIDVMVFTANPKKEPIVDVMNTVIS
AMALDYPTDKLAVYLADDGGCPLSLYAMEEACVFAKLWLPFCRKYGIKTRCPKAFFSPL
GDDERVLKNDDFDAEMKEIKLKYEEFQQNVERAGESGKINGNVVPDRASFIKVINDRKAE
SEKSADDLTKMPLLVYVSRERRFNRLHHFKGGSANALLRVSGIMSNAPYILVLDCDFFCH
DPISARKAMCFHLDPKLSSDLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWHGMDGIQGP
VLSGTGYFLKRKALYTSPGVKEEYLSSPEKHFGRSKKFLASLEEKNGYVKAEKVISEDIVE

EAKTLATCAYEDGTHWGQEIGYSYDCHLESTFTGYLLHCKGWRSTYLYPDRPSFLGCAP
VDMQGFSSQLIKWVAALTQAGLSHLNPITYGFSSRMKTLQCMCYAYLIYFTLYSWGMVL
YASVPSIGLLFGFQVYPDVHDPWFAVYVIAFISAILENMSESIPDGGSFKSWWMEYRALM
MMGVSAIWLGGLKAILDRIIGTEGEKLYLSDKAIDKEKLKKYEKGKFDFQGIGILAVPLIAF
SLLNLVGFIVGANHVFITMNYAGVLGQLLVSSFFVFVVVTVVIDVVSFLKVS [SEQ ID NO:
35]

[00845] Cellulose synthase like (*solanum chacoense*)

[00846] ATGAAAAAAACCATGGAGCTCAACAAAAGCACTGTTCCACAACCTATCACC
ACCATATACCGACTCCACATGTTCGTCCATTCTATAATCATGGCTGCATTAATATACTA
CCGTGTATCTAATTTGTTTAAATTCGAAAACATTCTGAGTTTACAAGCACTTGCTTGGG
TACTCATCACTTTTGGTGAATTTAGTTTCATTCTCAAGTGGTTCTTCGGACAAGGAACT
CGTTGGCGCCCTGTTGAAAGAGATGTTTTCCCTGAAAACATAACTTGCAAAGATTCCG
ATCTACCACCAATTGACGTAATGGTATTCACTGCCAATCCTAAGAAAGAGCCAATTGT
GGATGTCATGAACACTGTGATATCCGCAATGGCTCTAGATTATCCTACGGATAAATTG
GCTGTGTATCTGGCTGATGATGGAGGATGTCCTTTGTCATTGTACGCCATGGAAGAAG
CATGTGTGTTTGCAAAGCTGTGGCTACCTTTCTGTAGGAAGTATGGAATTAAAACCAG
ATGCCCTAAAGCGTTTTTTTCTCCTTTAGGAGATGATGACCGTGTTCTTAAGAATGATG
ATTTTGATGCTGAAATGAAAGAAATTAAATTGAAATATGAAGAGTTCCAGCAGAATG
TGGAACGTGCTGGTGAATCTGGAAAAATCAATGGTAACGTAGTGCCTGATAGAGCCT
CGTTTATTAAGGTAATAAACGACAGAAAAACGGAGAGCGAAAAGAGTGCCGATGATT
TAACGAAAATGCCCTTGCTAGTTTATGTATCCCGTGAAAGAAGATTCAACCGTCTTCA
TCACTTCAAGGGTGGATCTGCAAATGCTCTTCTTCGAGTTTCTGGAATAATGAGTAAT
GCCCCCTATATACTGGTGTTAGATTGTGATTTCTTCTGTCATGATCCAATATCAGCTAG
GAAGGCAATGTGTTTTCATCTTGATCCAAAGCTATCATCTGATTTAGCTTATGTTCAGT
TCCCTCAAGTCTTTTACAATGTCAGCAAGTCCGATATTTATGATGTCAAAATTAGACA
GGCTTACAAGACAATATGGCATGGAATGGATGGTATCCAAGGCCCAGTGTTATCAGG
AACTGGTTATTTTCTGAAGAGGAAGGCGTTATACACGAGTCCAGGAGTAAAGGAGGA
GTATCTTAGTTCACCGGAAAAGCATTTTGGAAGGAGTAAAAAGTTCCTTGCTTCACTA
GAGGAGAAAAATGGTTATGTTAAGGCAGAGAAAGTCATATCAGAAGATATCGTAGAG
GAAGCTAAGACCTTAGCTACTTGTGCATATGAGGATGGTACACATTGGGGTCAAGAG
ATCGGTTATTCATACGATTGTCATTTGGAGAGCACTTTTACTGGTTATCTATTACACTG
CAAAGGGTGGACATCGACTTATTTGTATCCAGACAGGCCATCTTTCTTGGGTTGTGCT
CCAGTTGATATGCAAGGTTTCTCCTCACAGCTCATAAAATGGGTTGCTGCACTTACAC
AAGCTGGTTTATCACATCTCAATCCCATCACTTATGGCTTGAGTAGCAGGATGAAAAC

TCTCCAATGCATGTGCTATGCCTATTTGATATATTTCACTCTTTATTCTTGGGGAATGG
TTCTATATGCTAGTATTCCTTCTATTGGTCTTTTGTTTGGCTTCCAAGTCTATCCGGAGG
TACATGATCCATGGTTTGCAGTGTATGTGATTGCTTTCATATCGACAATTTTGGAGAAT
ATGTCGGAGTCAATTCCAGAAGGGGGATCATTTAAATCGTGGTGGATGGAATACAGG
GCACTGATGATGATGGGAGTTAGTGCAATATGGTTAGGAGGATTGAAAGCTATATTA
GACAAGATAATCGGAACAGAAGGAGAGAAATTGTATTTGTCAGACAAGGCAATTGAC
AAGGAAAAGCTCAAGAAATACGAGAAGGGGAAATTTGATTTCCAAGGAATAGGGAT
ACTTGCTGTACCATTGATAGCATTTTCCCTGTTGAACCTGGTAGGCTTCATTGTTGGAG
CTAATCATGTCTTTATTACTATGAACTACGCAGGTGTGCTTGGCCAACTCCTCGTATCA
TCCTTCTTCGTCTTTGTCGTGGTCACTGTTGTCATTGATGTCGTTTCTTTCTTAAAGGTT
TCTTAA [SEQ ID NO: 36]

[00847] Cellulose synthase like (*solanum chacoense*)

[00848] MKKTMELNKSTVPQPITTIYRLHMFVHSIIMAALIYYRVSNLFKFENILSLQALA
WVLITFGEFSFILKWFFGQGTRWRPVERDVFPENITCKDSDLPPIDVMVFTANPKKEPIVDV
MNTVISAMALDYPTDKLAVYLADDGGCPLSLYAMEEACVFAKLWLPFCRKYGIKTRCPK
AFFSPLGDDDRVLKNDDFDAEMKEIKLKYEEFQQNVERAGESGKINGNVVPDRASFIKVI
NDRKTESEKSADDLTKMPLLVYVSRERRFNRLHHFKGGSANALLRVSGIMSNAPYILVLD
CDFFCHDPISARKAMCFHLDPKLSSDLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWHGM
DGIQGPVLSGTGYFLKRKALYTSPGVKEEYLSSPEKHFGRSKKFLASLEEKNGYVKAEKVI
SEDIVEEAKTLATCAYEDGTHWGQEIGYSYDCHLESTFTGYLLHCKGWTSTYLYPDRPSF
LGCAPVDMQGFSSQLIKWVAALTQAGLSHLNPITYGLSSRMKTLQCMCYAYLIYFTLYS
WGMVLYASIPSIGLLFGFQVYPEVHDPWFAVYVIAFISTILENMSESIPEGGSFKSWWMEY
RALMMMGVSAIWLGGLKAILDKIIGTEGEKLYLSDKAIDKEKLKKYEKGKFDFQGIGILA
VPLIAFSLLNLVGFIVGANHVFITMNYAGVLGQLLVSSFFVFVVVTVVIDVVSFLKVS
[SEQ ID NO: 37]

[00849] Cellulose synthase like (eggplant)

[00850] ATGAAAAAACAAATGGAGCTCAACAGAAGTGTTGTACCGCAACCTATCACC
ACCATTTACCGTCTCCACATGTTTATCCATGCCCTAATCATGCTAGCACTAATATACTA
CCGTGTCTCTAATTTGGCCAAATTCGAAAACATCCTCAGTTTACAAGCACTTGCTTGG
GCTCTTATCACGTTAGGTGAACTTTGTTTCATAGTCAAGTGGTTCTTCGGACAAGGGA
CTCGTTGGCGTCCTGTTGATAGGGATGTCTTCCCTGAAAACATCACTTGTCCAGATTCC
GAGCTACCCCCCATTGATGTCATGGTTTTCACTGCAAATCCTAAGAAAGAGCCAATTG
TGGATGTCATGAACACTGTCATATCCGCAATGGCTCTTGATTACCCGACCGACAAATT
GGCCGTTTATTTGTCTGATGATGGAGGATGCCCCTTGACGTTGTACGCAATGGAGGAA

GCTTGTTCCTTTGCCAAGTTGTGGCTACCTTTTTGTAGGAAGTATGGAATCAAAACAA
GGTGCCCTAAGGCGTTTTTTTCTCCATTAGGAGAAGATGACCGTGTATTGAAGAGTGA
TGACTTTGTTTCTGAAATGAAAGAAATGAAGTCAAAATATGAAGAGTTCCAGCAGAA
CGTGGACCGTGCTGGTGAATCCGGAAAAATCAAAGGTGACGTAGTGCCTGATAGACC
CGCGTTTCTTAAGGTACTAAATGACAGGAAGACGGAGAACGAGAAGAGTGCAGACGA
TTTAACTAAAATGCCTTTGCTAGTATACGTATCCCGTGAAAGAAGAACTCACCGTCGC
CATCACTTCAAGGGTGGATCTGCAAATGCTCTTCTTCGAGTTTCTGGGATAATCAGTA
ATGCCCCCTATATACTGGTTTTAGATTGTGATTTCTTCTGTCATGATCCAATATCAGCT
CGGAAGGCAATGTGTTTCCATCTTGATCCAAAACTATCACCTGACTTAGCTTACGTGC
AGTTCCCTCAAGTGTTTTACAATGTTAGCAAGTCCGATATTTACGACGTCAAAATTAG
ACAGGCTTACAAGACAATATGGCACGGGATGGATGGTATCCAAGGCCCAGTGTTATC
GGGAACTGGTTATTTTTTAAAAAAGAAGGCGTTGTACACGAGTCCAGGTCTAAAAGA
TGAGTATCTTAGTTCACCGGAAAAGCATTTCGGAACGAGTAGAAAGTTCATTGCTTCA
CTAGAGGAGAATAATTATGTTAAGCAAGAGAAAGTCATATCAGAAGATATCATAGAG
GAAGCTAAGAGACTGGCTACTTGTGCATACGAGGATGGCACACATTGGGGTCAAGAG
GCAAACAGGCCATCTTTCTTGGGTTGTGCCCCAGTTGATATGCAAGGTTTCTCCTCAC
AGCTCATAAAATGGGTTGCTGCACTCACACAAGCAGGTCTATCACATCTCAATCCCAT
CACTTACGGCTTCAAGAGCAGAATGAGAACTCTCCAAGTCTTGTGTTATGCCTATTTG
ATGTATTTCTCTCTTTATTCTTGGGGAATGGTTCTACATGCTAGTGTTCCTTCTATTGGC
CTTCTCTCTGGCATTAAAATCTACCCGGAGGTGTATGATCCATGGTTTGTTGTGTATGT
GATTGCTTTCATATCAACAATTTTGGAGAATATGTCGGAATCAATTCCGGAAGGGGGA
TCGGTTAAAACGTGGTGGATGGAATACAGGGCACTGATGATGATGGGAGTTAGTGCA
ATATGGCTAGGAGGAGTGAAAGCCATAGTAGACAAGATCATCGGAACGCAAGGAGA
GAAATTGTATTTGTCGGACAAAGCAATTGACAAGGAAAAGCTCAAGAAATACGAGAA
GGGGAAATTTGATTTCCAAGGAATAGGAATACTTGCTGTACCATTGATAACATTTTCT
GTGTTGAACCTGGTAGGCTTCTTGGTTGGAATTAATCAAGTGTTGATAACGATGAAGT
TCGCAGGCGTGCTGGGCCAACTCCTCGTATCATCCTTCTTCGTCTTTGTCGTCGTTACT
GTTGTCATTGATGTCGTATCTTTCTTAAAGGATTCTTAA [SEQ ID NO: 38]

[00851] Cellulose synthase like (eggplant)

[00852] MKKQMELNRSVVPQPITTIYRLHMFIHALIMLALIYYRVSNLAKFENILSLQALA
WALITLGELCFIVKWFFGQGTRWRPVDRDVFPENITCPDSELPPIDVMVFTANPKKEPIVD
VMNTVISAMALDYPTDKLAVYLSDDGGCPLTLYAMEEACSFAKLWLPFCRKYGIKTRCP
KAFFSPLGEDDRVLKSDDFVSEMKEMKSKYEEFQQNVDRAGESGKIKGDVVPDRPAFLK
VLNDRKTENEKSADDLTKMPLLVYVSRERRTHRRHHFKGGSANALLRVSGIISNAPYILV

LDCDFFCHDPISARKAMCFHLDPKLSPDLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWH
GMDGIQGPVLSGTGYFLKKKALYTSPGLKDEYLSSPEKHFGTSRKFIASLEENNYVKQEK
VISEDIIEEAKRLATCAYEDGTHWGQEANRPSFLGCAPVDMQGFSSQLIKWVAALTQAGL
SHLNPITYGFKSRMRTLQVLCYAYLMYFSLYSWGMVLHASVPSIGLLSGIKIYPEVYDPWF
VVYVIAFISTILENMSESIPEGGSVKTWWMEYRALMMMGVSAIWLGGVKAIVDKIIGTQG
EKLYLSDKAIDKEKLKKYEKGKFDFQGIGILAVPLITFSVLNLVGFLVGINQVLITMKFAGV
LGQLLVSSFFVFVVVTVVIDVVSFLKDS [SEQ ID NO: 39]

[00853] Cellulose synthase like (*capsicum annuum*)

[00854] ATGGAGCTCAACAGATGTACGGTGCAGCAACCTACCACTGCCATATACCGA
CTACACATGTTTCTCCACTCTCTAATCATGCTTGCATTAGTATACTATCGTTTGTCTAAT
CTGTTTTACTTCGAAAACGTCCTCACTTTACAAGCATTTGCATGGGGGCTTATCACCTT
AGGTGAAATTTGTTTCATTGTCAAGTGGTTCTTTGGTCAAGGGACTCGTTGGCGCCCC
GTTGTCAGGGAAGTGTTCCTGGACAATATTACTTGCCAAGATTCCGAGCTGCCCGCAC
TAGATGTGATGGTTTTCACTGCCAATCCCAAGAAAGAGCCAATTGTGGATGTCATGAA
CACTGTGATATCCGCAATGGCTCTTGATTACCCGACGGATAAATTGGCTGTGTATCTG
GCTGATGATGGAGGATGCCCCTTGACGTTGTACGCCATGGAGGAGGCCTGTTCTTTTG
CCAAGTTGTGGCTACCTTTCTGTAGGAAGTATGGAATCAAAACAAGGTGCCCCAAAG
CGTTTTTTTCTCCATTAGGAGAAGATGATCGTATCCTTAAGAACGATGACTTTGTAGCT
GAAATGAAAGAAATTAAATTAAAATATGAGGAGTTCCAGCAGAATGTAAACCTTGCT
GGTGAATCCGGAAAAATCAAAGGTGACGTAGTGCCTGATAGAGCCTCGTTTATTAAG
GTAATAAATGACAGGAAAATGGAGAACAAGAAGAGTGCCGACGATATAACGAAAAT
GCCTTTGCTAGTATACGTATCCCGTGAAAGAAGATTTAACAGTCGTCATCACTTCAAG
GGTGGATCTGCAAATGCTCTTCTTCGAGTTTCAGGGATAATGAGTAATGCCCCCTATT
TACTGGTCTTAGATTGTGATTTCTTCTGTCATGATCCAACATCAGCTCGGAAGGCAAT
GTGTTTCCATCTTGATCCAAAACTATCACCTTCCTTAGCTTATGTGCAGTTCCCTCAAG
TGTTTTACAATGTCAGCAAGTCCGATATATACGATGTCAAAATTAGACAGGCTTACAA
GACAATATGGCACGGAATGGATGGTATCCAAGGCCCAGTGTTATCGGGAACTGGGTA
TTTTCTGAAGAGGAAAGCGTTATACACGAGTCCAGGTCTAAAGGATGAGTATCTTATT
TCACCGGAAAAGCATTTCGGATCAAGTAGAAAGTTCATTGCTTCTCTAGAGGAGAAC
AATGGTTATGTTAAGCAAGAGAAACTCATAACAGAAGATATTATAGAGGAAGCGAAG
ACCTTGTCTACTTGTGCATACGAGGATGGTACACGATGGGGCGAAGAGATCGGTTATA
CCTACAATTGCCATTTGGAGAGCACTTTTACCGGCTATCTTTTGCACTGCAAAGGGTG
GACATCAACATATTTGTATCCAGAAAGGCCATCTTTCTTGGGTTGTGCCCCAGTTGAT
ATGCAAGGATTCTCCTCACAACTCACAAAATGGGTTGCTGCACTCACACAAGCTGGTC

TATCACATCTCAATCCCATCACTTACGGCATGAAGAGCAGGATTAAGACTATCCAATG
CTTGTGCTATGCCTATTTGATGTATTTCTCTCTCTATTCTTGGGGAATGGTTCTGCATGC
TAGTGTTCCTTCTATTAGCCTTTTGCTTGGCATTCAAGTCTACCCCGAGGTCTATGATC
CATGGTTTGCAGTGTATGTGCTTGCTTTCATATCGACAATTTTGGAGAACATGTCAGA
GTCAATTCCAGAAGGCGGTTCAGTTAAAACTTGGTGGATGGAATACAGGGCACTGAT
GATGATGGGAGTTAGTGCAATATGGTTAGGAGGAGTGAAAGCTATAGTAGAAAAGAT
CATCGGAACTCAAGGAGAGAAATTATATTTGTCGGACAAAGCAATTGACAAGGAAAA
GCTCAAGAAATATGAGAAGGGGAAATTTGATTTCCAAGGGATAGGGATACTTGCTGT
TCCATTGATAACATTCTCAGCGTTGAATTTGGTAGGCTTCATGGTTGGAGCTAATCAA
GTGATTCTTACTATGAAGTTCGAAGCTTTGCTAGGCCAACTCCTTGTGTCATCCTTCTT
CGTCTTTGTGGTGGTCACCGTTGTCATAGATGTCCTATCTTTCTTAAAAGACTCTTAA
[SEQ ID NO: 40]

[00855] Cellulose synthase like (*capsicum annuum*)

[00856] MELNRCTVQQPTTAIYRLHMFLHSLIMLALVYYRLSNLFYFENVLTLQAFAWG
LITLGEICFIVKWFFGQGTRWRPVVREVFLDNITCQDSELPALDVMVFTANPKKEPIVDVM
NTVISAMALDYPTDKLAVYLADDGGCPLTLYAMEEACSFAKLWLPFCRKYGIKTRCPKA
FFSPLGEDDRILKNDDFVAEMKEIKLKYEEFQQNVNLAGESGKIKGDVVPDRASFIKVIND
RKMENKKSADDITKMPLLVYVSRERRFNSRHHFKGGSANALLRVSGIMSNAPYLLVLDC
DFFCHDPTSARKAMCFHLDPKLSPSLAYVQFPQVFYNVSKSDIYDVKIRQAYKTIWHGMD
GIQGPVLSGTGYFLKRKALYTSPGLKDEYLISPEKHFGSSRKFIASLEENNGYVKQEKLITE
DIIEEAKTLSTCAYEDGTRWGEEIGYTYNCHLESTFTGYLLHCKGWTSTYLYPERPSFLGC
APVDMQGFSSQLTKWVAALTQAGLSHLNPITYGMKSRIKTIQCLCYAYLMYFSLYSWGM
VLHASVPSISLLLGIQVYPEVYDPWFAVYVLAFISTILENMSESIPEGGSVKTWWMEYRAL
MMMGVSAIWLGGVKAIVEKIIGTQGEKLYLSDKAIDKEKLKKYEKGKFDFQGIGILAVPLI
TFSALNLVGFMVGANQVILTMKFEALLGQLLVSSFFVFVVVTVVIDVLSFLKDS [SEQ ID
NO: 41]

[00857] The following sequences were generated for silencing GAME15 in their respective plants:

[00858] Region used for GAME15 silencing in Tomato

[00859] GGCTCTTGATTATCCCACCGATAAATTGGCTGTGTATCTCGCTGATGATGGA
GGATGTCCATTGTCGTTGTACGCCATGGAACAAGCGTGTTTGTTTGCAAAGCTATGGT
TACCTTTCTGTAGAAACTATGGAATTAAAACGAGATGCCCAAAAGCATTTTTTTCTCC
GTTAGGAGATGATGACCGTGTTCTTAAGAATGATGATTTTGCTGCTGAAATGAAAGAA
ATTAAATTGAAATATGAAGAGTTCCAGCAGAAGGTGGAACATGC [SEQ ID NO: 42]

[00860] Region used for GAME15 silencing in Potato

[00861] GGCTCTTGATTATCCTACGGATAAATTGGCTGTGTATCTGGCTGATGATGGA GGATGTCCTTTGTCATTGTACGCCATGGAAGAAGCATGTGTGTTTGCAAAGCTGTGGC TACCTTTCTGTAGGAAGTATGGAATTAAAACTAGATGCCCTAAAGCGTTTTTTTCTCCT TTAGGAGATGATGAACGTGTTCTTAAGAATGATGATTTTGATGCTGAAATGAAAGAA ATTAAATTGAAATATGAAGAGTTCCAGCAGAATGTGGAACGTGCTGGTG [SEQ ID NO: 43]

[00862] Region used for GAME15 silencing in Eggplant GGCTCTTGATTACCCGACCGACAAATTGGCCGTTTATTTGTCTGATGATGGAGGATGC CCCTTGACGTTGTACGCAATGGAGGAAGCTTGTTCCTTTGCCAAGTTGTGGCTACCTTT TTGTAGGAAGTATGGAATCAAAACAAGGTGCCCTAAGGCGTTTTTTTCTCCATTAGGA GAAGATGACCGTGTATTGAAGAGTGATGACTTTGTTTCTGAAATGAAAGAAATGAAG TCAAAATATGAAGAGTTCCAGCAGAACGTGGACCGTGCTGGTGAATCCGGAAAAATC AAAGGTGACGTAGTGCCTGATAGACCCGCGTTTCTTAAGGTACTAAATGACAGGAAG ACGGAGAACGAGAAGAGTGCAGACGATTTAACTAAAATGCCTTTGCTAGTATACGTA TCCCGTGAAAGAAGAACTCACCGTCGCCATCACTTCAAGGGTGG [SEQ ID NO: 44]

[00863] RNAi lines for the *GAME15* gene in tomato and potato were generated. *GAME15*-RNAi transgenic tomato plants showed severe reduction in α-tomatine and downstream SGAs in leaves; α-tomatine was not detected in *GAME15*-silenced green fruit. Furthermore, no esculeosides or other SGAs were detected during tomato fruit developmental stages (e.g., breaker and red fruit). In addition, a 15-20 fold increase in cholesterol, which is a precursor for SGAs was observed in leaves and green fruit of *GAME15*-RNAi tomato plants. In potato, silencing of *GAME15* resulted in a major reduction in α-chaconine and α-solanine, while the cholesterol pool in these lines increased.

Example 7: Generation of GAME15-RNAi transgenic tomato potato and eggplant plants

[00864] The *GAME15*-RNAi construct was generated by introducing a selected fragment (silencing sequences SEQ ID NO: 42 (tomato), SEQ ID NO: 43 (potato), and SEQ ID NO: 44 (eggplant)) to pENTR/D-TOPO (Invitrogen) (by *Not*I and *Asc*I) and further subcloning of this fragment to the pK7GWIWG2 (II) binary vector using the Gateway LR Clonase II enzyme mix (Invitrogen). The vector was transformed into tomato, potato and eggplant as described previously (Itkin et al. 2011. The Plant Cell 23:4507-25; Sonawane et al. 2018. PNAS 115(23): E5418-E5428). Positive *GAME25*-downregulated lines were further used for LC-MS analysis.

Example 8: GAME15-silenced tomato plants showed severely reduced SGA profile

[00865] In order to determine the precise role of GAME15 in SGA metabolism, GAME15-RNAi (GAME15i) transgenic tomato lines (#21, #22 and #23) were generated using the tomato silencing sequence above (SEQ ID NO: 42).

[00866] GAME15-RNAi leaves showed severe reduction in α-tomatine, compared with wild-type tomato leaves (**FIGURE 15A**). Furthermore, the SGAs profile of GAME15i fruit was subsequently compared to wild-type ones at different stages of development and ripening. During the transition from green to red fruit in tomato, α-tomatine is converted to esculeosides and lycoperosides, while dehydrotomatine is converted to dehydroesculeosides and dehydrolycoperosides (**FIGURE 14A**).

[00867] GAME15i green and red fruits did not show any trace of SGAs (e.g., α-tomatine or Esculeoside A) suggesting complete loss of SGAs in tomato fruits due to GAME15i silencing (**FIGURES 15B** and **15C**).

Example 9: Altering GAME15 expression has major impact on SGAs in potato

[00868] Similar to tomato, GAME15i was also silenced in potato (#1, #2, and #3) to determine its effect on potato SGAs metabolism, using the potato silencing sequence above (SEQ ID NO: 43).

[00869] Silencing of GAME15 in potato resulted in drastic reduction in α-chaconine (shaded bars) and α-solanine (open bars), major SGAs in potato leaf tissue (**FIGURE 16**), in comparison with potato leaf tissue of the wild-type.

Example 10: High cholesterol accumulation in GAME15-silenced tomato leaves

[00870] Cholesterol serves as a key precursor in the biosynthesis of SGAs (Sonawane et al., 2016, Nat. Plants 3: 16205). As severe reduction and subsequent complete loss of SGAs was observed in GAME15i-silenced tomato plants, the cholesterol levels in these plants were examined. An ~15-20-fold increase in cholesterol (SGA precursor) was observed in leaves of GAME15i-silenced tomato plants compared to the leaves of wild-type tomato plants (**FIGURE 17**).

Example 11: Altering GAME15 expression and observing its impact on SGAs in eggplant

[00871] Similar to potato, GAME15i is also silenced in eggplant to determine its effect on potato SGAs metabolism, using the eggplant silencing sequence above (SEQ ID NO: 44).

[00872] The effect of silencing of GAME15 in eggplant is observed with respect to reduced levels of α-solasonine and/or α-solamargine in comparison with wild-type eggplant (**FIGURE 14C**).

Example 12: Overexpression of GAME15 in tomato, potato, and eggplant

[00873] Alternatively, tomato, potato, and/or eggplant plants are genetically modified, or gene edited to overexpress GAME15.

[00874] To increase production of α-tomatine and esculeosides and/or lycoperosides in tomato plants (**FIGURE 14A**), tomato plants are genetically modified, or gene edited to overexpress GAME15.

[00875] To increase production of α-solanine and/or α-chaconine in potato plants (**FIGURE 14B**), potato plants are genetically modified, or gene edited to overexpress GAME15.

[00876] To increase production of α-solasonine and/or α-solamargine in eggplant (**FIGURE 14C**), eggplant plants are genetically modified, or gene edited to overexpress GAME15.

Example 13: Plants and crops with modified levels and compositions of SGAs

[00877] Based on the foregoing, *Solanaceous* plants (e.g., tomato, potato, eggplant, and/or pepper plants) and/or crops are prepared, such as through classical breeding or genetic engineering (e.g., genetically modified or transgenic plants, gene edited plants, and the like), with modified levels and compositions of SGAs, conferring on the plant a chemical barrier against a broad range of insects and other pathogens and/or removing anti-nutritional compounds (e.g., chaconine and/or solanine from potato).

[00878] Furthermore, high cholesterol or high phytosterol tomato lines are used to engineer high value steroidal compounds (e.g., pro-vitamin D and/or diosgenin), such as through synthetic biology tools.

[00879] In addition, high phytosterol (e.g., phytocholesterol) lines are used to produce components used in cosmetic products.

[00880] In other instances, Solanaceous plants (e.g., tomato, potato, eggplant, and/or pepper plants) and/or crops are prepared with increased levels of SGAs and/or decreased levels of phytosterols.

Example 14 - Materials and Methods for Examples 14-25

[00881] *Liquid Chromatography – Mass Spectroscopy (LC-MS) and Tandem Mass Spectroscopy (MS/MS) parameters for saponin analysis*

[00882] Four biological replicates (n=4) from sample were used for metabolic analysis. Briefly, 100 mg of frozen powdered plant tissue was extracted with 300 µL of 80% methanol mixed with internal standard (Ponasterone A, C = 1.5 µg/mL), briefly vortexed and then sonicated for 20 min. at room temperature. Extracts were centrifuged for 10 min. at 14,000 x *g* and filtered through 0.22 µm filters. Samples were analyzed using a high-resolution UPLC/qTOF system comprised of a UPLC (Waters Acquity) connected to a SYNAPT-G2 qTOF detector (tandem quadrupole/time-of-flight

mass spectrometer, Waters). Separation of metabolites was performed on a 100 x 2.1 mm i.d., 1.7 µm UPLC BEH C18 column (Waters Acquity). The mobile phase consisted of 0.1% formic acid in acetonitrile:water (5:95, v/v; phase A) and 0.1% formic acid in acetonitrile (phase B). The flow rate was 0.3 mL/min, and the column temperature was kept at 35°C.

[00883] The following linear gradient was used for analysis of triterpenoid saponins in spinach, *Beta vulgaris,* triterpenoids produced in yeast cells and *in vitro* with use of SOAP10: from 100% to 85% phase A over 5 min., from 85% to 75% phase A over 2 min., then held at 75% phase A for 3 min.; gradient continued to 65% A over 10 min., from 65% to 40% A over 2 min.; from 40% phase A to 100% phase B over 1 min., then held at 100% phase B 3.5 minutes and then returned to the initial conditions (100% phase A) within 0.5 min. and conditioning at 100% phase A for 1 min.

[00884] To analyze triterpenoid saponins from *Chenopodium quinoa* and *Medicago sativa* 40 min. gradient was used: from 100% to 72% phase A over 22 min., from 72% to 0% phase A over 14 min., then held at 100% phase B for 2 min.; and then returned to the initial conditions (100% phase A) within 0.5 min. and conditioning at 100% phase A for 1.5 min.

[00885] The following settings were used: capillary 2 kV; cone 27 V; source temperature was set to 140 °C, desolvation 450 °C, desolvation gas flow 800 L/h. Argon was used as the collision gas. Electrospray ionization (ESI) was used in negative ionization mode at the *m/z* range of 50-1600 Da. The mass spectroscopy (MS) system was calibrated using sodium formate, and Leu-enkephalin was used as the lock mass. MassLynx software version 4.1 (Waters) was used to control the instrument and calculate accurate masses and elemental compositions. In addition, a mixture of 15 standard metabolites, injected after each of the 10 samples, was used as quality controls. Data acquisition was performed in the MS^E mode with energy ramp that records an exact mass precursor and fragment ion information from every detectable component in a sample. MS^E mode rapidly alternates between two functions: the first acquiring low-energy exact mass precursor ion spectra and the second acquiring elevated energy exact mass fragment ion spectra. The collision energy for low-energy function was set to 4 eV, and for the high-energy to 15–50 eV ramp.

[00886] For separation of the products of the *Glycyrrhiza uralensis* enzymes in *N. benthamiana* another gradient was used: starting at 75% phase A for 5 min., then from 75% to 50% phase A over 15 min., from 50% to 30% phase A over 2 min., from 30% phase A to 100% phase B, then held at 100% phase B for 3.5 min. and then returned to the initial conditions (75% phase A) within 0.5 min. and conditioning at 75% phase A for 1 min. The flow rate was 0.3 mL/min., and the column temperature was kept at 35°C. Samples were analyzed using a high-resolution UPLC/qTOF system comprised of a UPLC (1290 Infinity II, Agilent) connected to an Impact HD UHR-QqTOF (Bruker). Separation of metabolites was performed on a 100 x 2.1 mm i.d., 1.7 µm UPLC BEH C18 column

(Waters Acquity). The mobile phase consisted of 0.1% formic acid in acetonitrile:water (5:95, v/v; phase A) and 0.1% formic acid in acetonitrile (phase B). ESI was used in negative ionization mode at the *m/z* range of 50-1700 Da with following parameters: drying gas: 200°C, 8 L/min; nebulizer: 2 Bar; capillary: 4.2 kV.

[00887] Metabolites were identified by comparing the retention times and mass fragments of standard compounds. When corresponding standards were not available, compounds were putatively identified by comparing their retention times, elemental composition and fragmentation pattern with those described in the literature (**Table 14**). Relative quantification of triterpenoid saponins in spinach was carried out using the TargetLynx (Waters) program. Peak area of saponins in each sample was normalized to peak area of internal standard in order to reduce variability originating from sample handling.

[00888] *Gas Chromatograph – Mass Spectroscopy (GC-MS) analysis of spinach triterpenoid aglycones*

[00889] Powdered frozen tissue (200 mg) was extracted 3 times with 600 µL of 80% MeOH. Collected fractions were evaporated in speed-vac (25˚C, 78 mbar, O/N). Dry residue was dissolved in 500 µL of 2M HCl in 50% MeOH and heated for 5 hours at 65˚C. Hydrolyzed samples were evaporated to dryness in the speed-vac (50˚C, 78 mbar, 1.5h). Dry residue was dissolved in 200 µL of Toluene: MeOH (3:2; v/v) and 75 µL of trimethylsilyldiazomethane ($TMSCHN_2$) were added. Methylation mixture was incubated at room temperature for 40 minutes, evaporated to dryness and resuspended in 80 µL of MSTFA, incubated for 20 min. at room temperature and 10 minutes at 65˚C. Sample was transferred to the glass insert and 1 µL was injected onto GC-MS as described before (*31*). The GC-MS system comprised a COMBI PAL autosampler (CTC Analytics), a trace GC ultra-gas chromatograph equipped with a programmable temperature vaporizing (PTV) injector, and a DSQ quadrupole mass spectrometer (Thermo Electron). GC was performed on a 30 m × 0.25 mm × 0.25 µm Zebron ZB-5 ms MS column (Phenomenex). The PTV split technique was performed as follows: Samples were analyzed in the constant temperature splitless mode. PTV inlet temperature was set at 280°C. Analytes were separated using the following chromatographic conditions: Helium was used as carrier gas at a flow rate of 1.2 mL/min. The thermal gradient started at 170°C, was held at this temperature for 1.5 min, ramped to 280°C at 37°C/min. and then ramped to 300°C at 1.5°C/min. and held at 300°C for 5.0 min. Eluents were fragmented in the electron impact mode with an ionization voltage of 70 eV. The reconstructed ion chromatograms and mass spectra were evaluated using Xcalibur software (ThermoFinnigan). Compounds were identified by comparison of their retention index and mass spectrum to those generated for authentic standards analyzed on the same instrument and those reported in literature (E. Biazzi *et al.*, CYP72A67 Catalyzes a Key

Oxidative Step in *Medicago truncatula* Hemolytic Saponin Biosynthesis. *Molecular Plant.* 8, 1493–1506 (2015)).

[00890] *Saponin purification for NMR analysis*

[00891] The system consisted of an Agilent 1290 Infinity II UPLC system equipped with a quaternary pump, auto sampler, diode array detector, a Bruker/Spark Prospekt II LC-SPE system (Spark) and Impact HD UHR-QqTOF mass spectrometer (Bruker) connected via a Bruker NMR MS Interface (BNMI-HP) as described previously (B. Khakimov, L. H. Tseng, M. Godejohann, S. Bak, S. B. Engelsen, Screening for Triterpenoid Saponins in Plants Using Hyphenated Analytical Platforms. *Molecules.* 21, pii: E1614 (2016)). MS spectra, in negative mode, were acquired between *m/z* 50 and 1700. The calibration was done with a 10 mM sodium trifluoroacetate (NaTFA, Sigma-Aldrich) automatically which is introduced at the beginning and the end of each chromatographic run. Separation was done on a XBridge LC column (BEH C18, 5 µm particle size, and 250 mm × 4.6 mm; Waters). The chromatographic conditions were as following; a flow rate of 0.9 mL/min. starting with a solvent composition of 74% A (5% ACN + 0.1% FA) and 26% B (100% ACN + 0.1% FA) with a linear gradient to 70% A at 60 min., followed by another linear gradient to 100% B at 62 min. 100% B hold for 3 min. followed linear gradient to 74% A at 65.5 and hold for 4.5 min. for equilibration. Saponins were collected on SPE cartridges in preset time windows and trapping was triggered by intensity threshold. Yossoside IV (*m/z* 1263.58, threshold 200'000, time window 18-23 min.); Yossoside XII (*m/z* 1233.57, threshold 58'000, time window 27-32 min.), Yossoside Va (*m/z* 1305.59, threshold 200'000, time window 32-40 min.); Yossoside X (*m/z* 1437.64, threshold 150'000, time window 40-48 min.) and Yossoside V (*m/z* 1305.59, threshold 200'000, time window 54-65 min.). For this trapping process, a makeup-flow of 2.5 mL/min. water was added to the eluent before it passed through the SPE cartridges in order to increase the retention of analytes on the cartridges. For the trapping 10 mm × 2 mm SPE cartridges filled with GP resin were used. Each cartridge was loaded five times with the same compounds, 5-10 cartridges were used for trapping one saponin. Prior to NMR measurements, SPE cartridges were dried with a stream of nitrogen, the fraction from each cartridge was eluted witha a total of 150 µL deuterated methanol. The sample was eluted into 96 well plate. Eluents containing same compound were pooled, dried under stream of nitrogen, freeze-dried, resuspended in 200 µL of D_2O and freeze-dried again to remove traces of H_2O. Dry compounds were dissolved in 60 µL 90% MeOD-*d4* 10% D_2O with 0.01% addition of 3-propionic-2,2,3,3-*d4* acid sodium salt (TMSP, that was used as an internal chemical shift reference for ^1H and ^{13}C spectra) and transferred to a 1.7 mm NMR test tube.

[00892] *NMR methods*

[00893] NMR spectra were recorded on a Bruker AVANCE NEO-600 NMR spectrometer

equipped with a 1.7 mm TXI-z and a 5 mm TCI-xyz CryoProbes equipped with shielded gradient coils. All spectra were acquired at 293 K. NMR data of Yossoside V were recorded on both the 1.7 and 5 mm CryoProbes, and spectra of medicagenic acid 3-*O*-glucuronide (MA-3-GlcA) were acquired on the 5 mm CryoProbe (using a special NMR spinner turbine to hold the 1.7 mm test tube in a 5 mm probe).

[00894] ^{1}H and ^{13}C chemical shift assignment was based on different 1D and 2D NMR techniques, all using pulsed field gradient selection.

[00895] The structure and stereochemistry of the aglycone unit were determined by 1D ^{1}H and ^{13}C, 2D homonuclear COSY and 2D heteronuclear HSQC and HMBC spectra. Identification of the monosaccharide units required recording also 2D homonuclear TOCSY spectra so as to complete the assignment and delineate connectivities between ^{1}H signals within the individual sugar units. Correlations observed in HMBC spectra between ^{1}H and ^{13}C that are 2-3 bonds apart, are shown in **FIGURE 21** and **FIGURE 35D** for Yossoside V and MA-3-GlcA, respectively.

[00896] ^{1}H 1D NMR spectra were acquired using 16 k data points and a recycling delay of 2.5 s. ^{13}C spectral-editing DEPTQ NMR spectra were acquired using 65 k data points and a recycling delay of 3 s.

[00897] 2D ^{1}H-^{1}H COSY, TOCSY and ROESY spectra were acquired using 16384-8192 (t_2) x 400-512 (t_1) data points. 2D TOCSY spectra were acquired using isotropic mixing times of 100-300 ms. A T-ROESY experiment was used in this study, TOCSY-less ROESY that effectively suppresses TOCSY transfer in ROESY experiments. 2D T-ROESY spectra were recorded using spin lock pulses of 100-400 ms.

[00898] 2D ^{1}H-^{13}C HSQC and HMBC spectra were recorded using 4096 (t_2) x 400-512 (t_1) data points. Multiplicity editing HSQC enables differentiating between methyl and methine groups that give rise to positive correlation, vs methylene groups that appear as negative peaks. HMBC delay for evolution of long range couplings was set to observe long range couplings of $J_{H,C} = 8$ Hz.

[00899] Chemical shift assignment was based on combined information derived from HMBC connectivities (**FIGURE 21** and **FIGURE 35D**) and TOCSY correlations (data not shown).

[00900] ^{1}H and ^{13}C chemical shift values of Yossoside V and MA-3-GlcA are shown in **Table 1**. As expected, chemical shift values of the aglycone unit of the two compounds nicely fit and similar HMBC correlation pattern is observed (**FIGURE 21** and **FIGURE 35D**). Assignment of the protons as axial or equatorial was based on the observed vicinal *J* couplings; large value (>10 Hz) indicates on axial protons, further supported by correlations observed in ROESY spectra.

[00901] HMBC correlations in aglycone region are in accordance with those observed in other triterpenoid saponins (H. Schröder *et al.*, A triterpene saponin from Herniaria glabra. *Phytochemistry*.

34,1609-13 (1993); E. P. Mazzola *et al.*, Utility of coupled-HSQC experiments in the intact structural elucidation of three complex saponins from Blighia sapida. *Carbohydr Res.* 346, 759-68 (2011)). The linkage site between medicagenic acid and glucuronic acid in MA-3-GlcA was supported by strong HMBC connectivity between the anomeric hydrogen of GlcA and C3 of the aglycone, as well as a complementary correlation observed between MA-H3 / anomeric C1 of GlcA (red arrow in **FIGURE 18D**). Similarly, linkage sites between the saccharide units were determined based on complementary 3-bond HMBC connectivities observed between the anomeric hydrogen/carbon of one sugar and the proton/carbon of the directly attached glycan as indicated by the interglycan arrows between Fuc-Rha-Glc (**FIGURE 21**). The addition of xylose in Yossoside V (Xyl connected to GlcA) seems to restrict the conformational space of GlcA as is indicated by a lack of previously observed correlation between GlcA and the aglycone and lack of *J* couplings within GlcA. However, chemical shift prediction further supports the assignment of GlcA in Yossoside V.

[00902] *RNA-Seq library preparation and data analysis*

[00903] RNA-Seq libraries were prepared as described previously (S. Zhong *et al.*, High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc.* **8**, 940-9 (2011)) with minor modifications. Briefly, 5 µg of total RNA was used for poly(A) RNA capture using Dynabeads Oligo $(dT)_{25}$ (Invitrogen), fragmented at 94 °C for 5 minutes and eluted. The first-strand cDNA was synthesized using reverse transcriptase SuperScript III (Invitrogen) with random primers and dNTP, whereas the second-strand cDNA was generated using DNA polymerase I (Enzymatics) using dUTP. After end-repair (Enzymatics), dA-tailing with Klenow 3′-5′ (Enzymatics) and adapter ligation (Quick T4 DNA Ligase, NEB), the dUTP-containing second-strand was digested by uracil DNA glycosylase (Enzymatics). The resulting first-strand adaptor-ligated cDNA was used for PCR enrichment (NEBNext High-Fidelity PCR Master Mix, NEB) for 14 cycles. Indexed libraries were pooled and sequenced (paired-end, 125 bp) on a single lane of HiSeq2500 (Illumina) at the Crown Institute for Genomics, The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science. The quality of the raw data was assessed using FastQC (S. W. Wingett, S. Andrews, FastQ Screen: A tool for multi-genome mapping and quality control. Version 2. *F1000Res.* **7,** 1338 (2018)). Raw reads were aligned against version 1.0.1 of the spinach reference genome (The *Beta vulgaris* resource (http://bvseq.boku.ac.at/Genome/Download/index.shtml)) using Tophat v2.0.13 (C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 25, 1105-11 (2009)). Quantification was performed using HTSeq (S. Anders S, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 31, 166-9 (2015)) followed by normalization and differential gene expression analysis using DESeq2

(M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014)).

[00904] *Co-expression analysis*

[00905] Co-expression analysis was done using CoExpNetViz software (O. Tzfadia, T. Diels, S. De Meyer, K. Vandepoele, A. Aharoni,Y. Van de Peer, CoExpNetViz: Comparative Co-Expression Networks Construction and Visualization Tool. Front Plant Sci. 6, 1194 (2016)). Briefly, spinach *bAS* (*SOAP1*), *CYP716A268* (*SOAP2*) and *CYP716A268v2* (*SOAP2-like*) were used as 'baits' in co-expression analysis with correlation threshold set at 5 and 95 for lower ad higher percentile rank, respectively. Lists of co-expressed genes was additionally filtered according to correlation coefficient ($r > 0.9$; PCC) for each bait (**Table 14**). The analysis was performed using spinach RNA-Seq transcriptome data from different tissues and developmental stages (SO4WOLLD – mature leaf from four week old plant grown in long day; SO8WOLSL – mature leaf from eight week old plant grown in short day; SO4WYLLD – young leaf from four week old plant grown in long day; SO8WOLSL – young leaf from eight week old plant grown in short day; SO4WFBLD – flower bud from four week old plant grown in long day). The co-expression network was visualized with the Cytoscape software (P. Shannon P et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498-504 (2003)).

[00906] *Cloning*

[00907] Phusion High-Fidelity DNA Polymerase (New England Biolabs) was used for all PCR amplification steps according to the manufacturer's instructions. Restriction enzymes and T4 Ligase used for cloning were purchased from New England Biolabs. Oligonucleotide primers were purchased from Sigma-Aldrich. DNA excised from agarose gels was purified using the Gel/PCR extraction kit (Hy-Labs). *E. coli* TOP10 cells (Invitrogen) were used for plasmid isolation prior to transformation into other heterologous hosts. Plasmid DNA was isolated from *E. coli* cultures using the AccuPrep® Plasmid Mini Extraction Kit (Bioneer). For a list of primers used for cloning, see **Table 11**.

[00908] *Construct preparation for VIGS*

[00909] For the silencing of candidate genes in spinach, all gene sequences (200-400 bp) were amplified from spinach leaf cDNA template. Purified amplicons were inserted into modified pTRV2 vector (~200bp of magnesium chelatase subunit H inserted with EcoRI) digested with SacI.

[00910] *GoldenBraid system and expression in N. benthamiana*

[00911] For *N. benthamiana* transient expression of multiple genes from the same Agrobacterium strain, multigene constructs were created using GoldenBraid 3.0 (GB) system (M. Vazquez-Vilar et al., GB3.0: a platform for plant bio-design that connects functional DNA elements with associated

biological data. Nucleic Acids Res. 45, 2196-2209 (2017)). In the first step, all the CDS and tomato Ubiquitin promoter and terminator were domesticated by removing BsaI and BsmBI restriction sites and inserting them into pUPD2 GoldenBraid entry vector. In the next step, genes were subcloned together with terminator and promoter into pDGB3α1 or pDGB3α2 GB vector. First level α vectors allow transient expression of a single gene in *N. benthamiana*. Subsequently, pDGB3α1:SOAP1 was combined with pDGB3α2:SOAP2 into pDGB3Ω1:SOAP1+SOAP2 and pDGB3α1:SOAP3 with pDGB3α2:SOAP4 into pDGB3Ω2:SOAP3+SOAP4. Combination of the omega vectors allows expression of four genes providing production of medicagenic acid in *N. benthamiana*. Other genes, *cellulose synthase like G* (*SOAP5*), glycosyltransferases (*SOAP6-9*) and acyltransferase (*SOAP10*) were used in alpha vectors. Genes from *C. quinoa*, *B. vulgaris*, *L. japonicus*, *Glycine max* (*G. max*,) *M. sativa* and *G. uralensis* were cloned using same approach. For cDNA template preparation RNA extracted from 2-week-old seedlings (*L. japonicus*, *G. max*, *M. sativa* and *G. uralensis*), mature leaves (*B. vulgaris*) or from leaves surrounding flower buds (*C. quinoa*) was used.

[00912] *Cloning into yeast system (pESC)*

[00913] For the expression of *SOAP1* (*bAS*), CYPs (*SOAP2-4*) and *CSLG* (*SOAP5*) in *S. cerevisiae*, the sequences were amplified from cloned pDGB3α vectors carrying the sequence of interest and the purified amplicons were inserted into series of pESC (Amp^R) plasmids allowing simultaneous expression of two genes from one plasmid. *SOAP1* and *SOAP2* were inserted into pESC-HIS plasmid using NotI/SacI and BamHI/SalI restriction enzymes, respectively. *SOAP3* and *SOAP4* were inserted into pESC-LEU plasmid using NotI/SacI and BamHI/XhoI restriction enzymes, respectively. *SOAP5* and *SoUGD1* were inserted into pESC-LEU plasmid using NotI/SacI and SalI/XhoI restriction enzymes, respectively.

[00914] *Cloning into pET-28b and expression in E coli*

[00915] *SOAP10* was cloned into the pET-28b (Kan^R) vector using NdeI/NotI restriction sites and expressed in *E. coli* BL21 (DE3) cells.

[00916] In vitro enzymatic assay with SOAP10

[00917] Bacterial cells transformed with pET-28b:SOAP10 were grown in LB medium at 37°C. When cultures reached A600 = 0.6, protein expression was induced with 200 μM of isopropyl-1-thio-*β*-D-galactopyranoside (IPTG) at 15°C, for 24 h. Bacterial cells were lysed by sonication in 50 mM Tris-HCl pH 8.0, 10% glycerol, 0.1% Triton X-100, 1 mM PMSF and 100mg/mL lysozyme.

[00918] The fraction of spinach desacetyl saponins was used as a substrate in SOAP10 enzyme assay. In a 200 μL tube 6 μL of cell lysate was mixed with 2 μL of 6.18 mM acetyl coenzyme A (Acetyl-CoA) and 2 mL of saponin substrate (10 μg). Reaction was carried out for 2.5h at 28°C. Cell lysate without SOAP10 was used as a control. The reactions were stopped by addition of 20 μL

methanol, followed by brief vortex. Finally, the extracts were centrifuged for 10 min. at 14,000 x *g* and analyzed by LC-MS, as described above.

[00919] *VIGS Procedure*

[00920] Virus Induced Gene Silencing was performed as described previously (M. Senthil-Kumar, K. S. Mysore, Tobacco rattle virus-based virus-induced gene silencing in *Nicotiana benthamiana.* Nat Protoc. 9, 1549-62 (2014)). Briefly, pTRV1 and pTRV2 constructs harboring sequence for silencing were transformed into *Agrobacterium tumefaciens* (GV3101) electrocompetent cells. Transformants were grown on LB plates containing 50 µg/mL kanamycin and 50 µg/mL gentamicin at 28°C. 10 mL of LB medium supplemented with antibiotics was inoculated with a single colony and grew O/N at 28°C. Cells were centrifuged at 3,000 x *g* for 10 min. and supernatant removed. Pellet was resuspended in 5 mL of infiltration medium (100 mM MES buffer, 2mM $Na_3PO_4 \cdot 12H_2O$, 100 µM acetosyringone) and centrifuged again. Pellet was resuspended again in 10 mL of infiltration medium and incubated at room temperature for 2h. *Agrobacterium* suspensions (OD600 = 0.3 for each strain) were infiltrated into the underside of 1-week old spinach cotyledons (one per plant) with a needleless 1 mL syringe. Plants were grown in 16h photoperiod at 23°C. After 2 weeks first signs of silencing of *Magnesium Chelatase subunit H* (CHCL) were visible. Plants were grown for another 2 weeks prior tissue collection. Leaves were harvested, frozen in liquid nitrogen and stored at -80°C for later processing. Biological replicates consisted of 4 leaves all from different plants. **Table 10** below provides the spinach and red beet nucleic acid sequences used to silence CSLGs in red beet and spinach

[00921] **Table 10: VIGs Sequences**

Name	SEQ ID NO:	Nucleic Acid Sequence
SoCSLG (SOAP5) VIGS spinach	106	TTCCGTAGCCATCTTCTCGCTCTTCTACTACCGTTTCACTTCCTTCTTCA ACTCCGACATCTCCATACTTGCTTACTCCTTACTCACCACCGCCGAACT CTTCTTAACCTTTCTATGGGCTTTTACTCAGGCTTTCCGGTGGCGTCCCG TAATGAGGGAAGTCTCCGGGTACGAATCCATCAAACCCGAACAACTAC CGGGTTTGGATGTCTTCATTGTCACTGCTGACCCGACAAAGGAGCCAGT TCTGGAGGTGATGAACTCCGTGATATCATCCATGGCGTTGGATTATCCG GTTGATAGACTGGCGGTTTACTTGTCGGATGACGGTGGTTCTCCGTTGT CGAAGG
BvCSLG VIGS_ red beet	107	CCTTCGGCTCCTCCACAAAATTCATTGCGTCAGTTAGTTCAAACTCCAA GCAAAATATGGCCTTGAAGGAAATGACAAGAGACGACTTGTTAGAAGA GGCTAAAAATTTGGCTACTTGTGCATATGAATCAAACACTGAATGGGG TAACAAGATTGGATATTCGTATGAGTGTTTGTTGGAGAGTACATTTACC GGATATCTCTTACATTGCAAAGGATGGATTTCCGTGTATCTTTACCCAA AAAGACCCTGCTTCTTAGGATGCACGACGATTGACATGAAAGATGCCA TGGTTCAACTAATGAAATGGACCTCTGGATTACTAGGAGTTGGCATATC AAAGTTTAGCCCTCTAACTTATGCCTTTTCGAGGATGTCTATATTACAA AGCATGTGCTACGGTTACTTCACATTTTCAGCCCTTTTCGGAGTTTCGT

[00922] *Site-directed mutagenesis of SOAP5*

[00923] Site-directed mutagenesis of SOAP5 was performed using QuikChange Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions.

[00924] *Gene expression analysis – qPCR*

[00925] Gene expression analysis was performed with three/four biological replicates (n=3/4) for each VIGS silenced line. RNA isolation was performed by the Trizol method (Sigma-Aldrich). DNase I (Sigma-Aldrich)-treated RNA was reverse transcribed using a high-capacity cDNA reverse transcription kit (Applied Biosystems). Gene-specific oligonucleotides were designed with Primer Blast software (NCBI). The *Translation Elongation Factor Alpha 1* (*EF1-α*) gene was used as an endogenous control. Oligonucleotides used are listed in **Table 11**.

[00926] **Table 11: List of primers used**

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
2	**OXIDOSQUALENE CYCLASE**						
3	SOAP1	β-amyrin synthase, *bAS*	*Spinacia oleracea*	Sp_107620_kqnh	VIGS	Sp107620_vF	CTAGAGCTCATGTTCAAGAAGTTAT ACCC
4						Sp107620_vR	TTCGAGCTCCCTCCTCTTTTGTGTAGT AA
5					expression in *N. benthamiana*	Sp107620_GB_1F	GCGCCGTCTCGCTCGAATGTGGAG GTTGAAGGTTGG
6						Sp107620_GB_1R	GCGCCGTCTCGCTGTCTCATATGTG ATCTCCT
7						Sp107620_GB_2F	GCGCCGTCTCGACAGCCTCGACGA CATTAAAG
8						Sp107620_GB_2R	GCGCCGTCTCGCTCAAAGCTCAAGT AGAGTTGATAGAAGGTAA
9					expression in yeast	Sp107620_NotI_F	AAAAGCGGCCGCATGTGGAGGTTG AAGGTTGGA
10						Sp107620_SacI_R	AAAAGAGCTCTCAAGTAGAGTTGAT AGAAGGTAATG
11					colocalization	Sp107620_GB_1F	GCGCCGTCTCGCTCGAATGTGGAG GTTGAAGGTTGG
12						Sp107620_GB_CT_R	GCGCCGTCTCGCTCACGAACCAGT AGAGTTGATAGAAGGTAATGA
13	**CYTOCHROMES P450**						
14	*SoCYP1*	*CYP715A56*	*Spinacia oleracea*	Sp_046690_sjw	VIGS	Sp046690_vF	CTAGAGCTCCTTCGTCGTTGTTGTC GTCG
15						Sp046690_vR	TTCGAGCTCCCTATGACGAGCCCAA GCAT
16	*SoCYP2*	*CYP72A656*	*Spinacia oleracea*	Sp_148230_dgra	VIGS	Sp148230_vF	CTAGAGCTCAGCTCTACTTTGCGTG CTCA
17						Sp148230_vR	TTCGAGCTCTCTGGCCCTCATAGTT CGGA
18	*SoCYP3*	*CYP72A685*	*Spinacia oleracea*	Sp_085350_pxis	VIGS	Sp085350_vF	CTAGAGCTCTTGTTGGGTTGCTTGT CCCT
19						Sp085350_vR	TTCGAGCTCGGTCTTTGTTGGCACG AACC
20	*SOAP2*	*CYP716A268*	*Spinacia oleracea*	Sp_107660_kiqg	VIGS	Sp107660_vF	CTAGAGCTCTGGTGGGCGAGTCATT TGAA
21						Sp107660_vR	TTCGAGCTCCAAGCCAGCCAGAAG GTGTA

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
22					expression in *N. benthamiana*	Sp107660_GB_1F	GCGCCGTCTCGCTCGAATGGAACT CTTCTTTATGTGTGG
23						Sp107660_GB_1R	GCGCCGTCTCGCTCAAAGCTTAAGC AGCTACAGTTCGAGG
24					expression in yeast	Sp107660_BamHI_F	AAAAGGATCCATGGAACTCTTCTTTA TGTGTGGG
25						Sp107660_SalI_R	AAAAGTCGACTTAAGCAGCTACAGT TCGAGGATG
26					colocalization	Sp107660_GB_1F	GCGCCGTCTCGCTCGAATGGAACT CTTCTTTATGTGTGG
27						Sp107660_GB_CT_R	GCGCCGTCTCGCTCACGAACCAGC AGCTACAGTTCGAGGAT
28	*SOAP3*	*CYP72A655*	*Spinacia oleracea*	Sp_085340_meek	VIGS	Sp085340_vF	CTAGAGCTCGTTGCATTGCTTGTCC CTGG
29						Sp085340_vR	TTCGAGCTCTGATTGCAGAAGGCGT AGGG
30					expression in *N. benthamiana*	Sp085340_GB_1F	GCGCCGTCTCGCTCGAATGATAGAA ATCGGGTATATTGTAAA
31						Sp085340_GB_1R	GCGCCGTCTCGCTCAAAGCTTAGTC CCTGAGCTTATGTATAAT
32					expression in yeast	Sp085340_NotI_F	AAAAGCGGCCGCATGATAGAAATCG GGTATATTGTAAAATG
33						Sp085340_SacI_R	AAAAGAGCTCTTAGTCCCTGAGCTT ATGTATAATGAG
34					colocalization	Sp085340_GB_1F	GCGCCGTCTCGCTCGAATGATAGAA ATCGGGTATATTGTAAA
35						Sp085340_GB_CT_R	GCGCCGTCTCGCTCACGAACCGTC CCTGAGCTTATGTATAATG
36	*SOAP4*	*CYP72A654*	*Spinacia oleracea*	Sp_040350_wdny	VIGS	Sp040350_vF	CTAGAGCTCTTTCAAAGAGCGCGAG TGGT
37						Sp040350_vR	TTCGAGCTCGCCTGCACCAGAATTG TTGA
38					expression in *N. benthamiana*	Sp040350_GB_1F	GCGCCGTCTCGCTCGAATGATTTCA AAGAGCGCGAGT
39						Sp040350_GB_1R	GCGCCGTCTCGTTGTCTCTTGACCA GCCAAG
40						Sp040350_GB_2F	GCGCCGTCTCGACAACCTCAGTGG CCCTAAC

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
41						Sp040350_GB_2R	GCGCCGTCTCGCTCAAAGCTTAAAA TCGATGTAAAATAATGTGGG
42					expression in yeast	Sp040350_BamHI_F	AAAAGGATCCATGATTTCAAAGAGC GCGAGT
43						Sp040350_XhoI_R	AAAACTCGAGTTAAAATCGATGTAA AATAATGTGGG
44					colocalization	Sp040350_GB_1F	GCGCCGTCTCGCTCGAATGATTTCA AAGAGCGCGAGT
45						Sp040350_GB_CT_R	GCGCCGTCTCGCTCACGAACCAAAT CGATGTAAAATAATGTGGGC
46	GuCYP88D6	*CYP88D6*	*Glycyrrhiza uralensis*	Glyur000561s0002345 1.1	expression in *N. benthamiana*	GuCYP88D6_GB_1F	GCGCCGTCTCGCTCGAATGGAAGT ACATTGGGTTTGC
47						GuCYP88D6_GB_1R	GCGCCGTCTCGCTCAAAGCCTAAG CACATGATACCTTTATCAC
48	GuCYP72A154	*CYP72A154*	*Glycyrrhiza uralensis*	Glyur000890s0001907 1.1	expression in *N. benthamiana*	GuCYP72A154_GB_1F	GCGCCGTCTCGCTCGAATGGATGC ATCTTCCACACC
49						GuCYP72A154_GB_1R	GCGCCGTCTCGAAAGACCAATGGAT TTTGATTGTG
50						GuCYP72A154_GB_2F	GCGCCGTCTCGCTTTCCAAAGATGA TGCTGCA
51						GuCYP72A154_GB_2R	GCGCCGTCTCGTTGTCTCTTGCCCT GCCAGG
52						GuCYP72A154_GB_3F	GCGCCGTCTCGACAACCGCAGCTTT GCTGGC
53						GuCYP72A154_GB_3R	GCGCCGTCTCGCTCAAAGCTTACAG TTTATGCAGAATGATGGG
54	**GLYCOSYLTRANSFERASES**						
55	*SOAP5*	*Cellulose Synthase Like G2, CSLG2*	*Spinacia oleracea*	Sp_076690_ejcm/Spo12715	VIGS	Sp076690_vF	CTAGAGCTCTTCCGTAGCCATCTTC TCGC
56						Sp076690_vR	TTCGAGCTCCCTTCGACAACGGAGA ACCA
57					expression in *N. benthamiana*	Sp076690_GB_1F	GCGCCGTCTCGCTCGAATGGCAAC TTCTCACATTCGC
58						Sp076690_GB_1R	GCGCCGTCTCGGAAGACGGTTAAC AATGGCTC
59						Sp076690_GB_2F	GCGCCGTCTCGCTTCACATCTTCCT CCATTCC

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
60						Sp076690_GB_2R	GCGCCGTCTCGGCCTCTTTTCCCTG GCTACA
61						Sp076690_GB_3F	GCGCCGTCTCGAGGCCTGGTCGTC CTCATCG
62						Sp076690_GB_3R	GCGCCGTCTCGCTCAAAGCTTATAA CCATCCCTTAACAACAGG
63					expression in yeast	Spo12715F_NotI_F	AAAAGCGGCCGCATGGCAACTTCTC ACATTCGCAA
64						Spo12715R_SacI_R	AAAAGAGCTCTTATAACCATCCCTTA ACAACAGGGTAG
65					colocalization	Sp076690_GB_1F	GCGCCGTCTCGCTCGAATGGCAAC TTCTCACATTCGC
66						Sp076690_GB_CT_R	GCGCCGTCTCGCTCACGAACCTAAC CATCCCTTAACAACAGG
67					mutagenesis	Spo12715_M1_F	CGTATGAGTGCTTGTTGGAGGATAC ATTCACTGGATATATG
68						Spo12715_M1_R	CATATATCCAGTGAATGTATCCTCCA ACAAGCACTCATACG
69						Spo12715_M2_F	CTACGGTTCAACTAATAAGATGGAC CTCCTCATTACTTGG
70						Spo12715_M2_R	CCAAGTAATGAGGAGGTCCATCTTA TTAGTTGAACCGTAG
71	*SoGT1*	*UGT79M1*	*Spinacia oleracea*	Sp_049940_mmut	VIGS	Sp049940_vF	CTAGAGCTCAGCTTTCCTCCCATGG CATC
72						Sp049940_vR	TTCGAGCTCACTTGTGGCTGGAGTT GAGG
73	*SoGT2*	*UGT79N1*	*Spinacia oleracea*	Sp_148240_fnwi	VIGS	Sp148240_vF	CTAGAGCTCCCCTCCTCGGCTATAC ACCT
74						Sp148240_vR	TTCGAGCTCCCTACGTCTGCCACAG TCTC
75	*SOAP6*	*UGT74BB2*	*Spinacia oleracea*	Sp_170930_hjgq	VIGS	Sp170930vF	CTAGAGCTCCCGAAGCCTCACTCCA ACTT
76						Sp170930vR	TTCGAGCTCTCAGAACTCGGGGATT GTGC
77					expression in *N. benthamiana*	Sp170930_GB_1F	GCGCCGTCTCGGCTCGAATGACGGG AAAAGGAAGAACG
78						Sp170930_GB_1R	GCGCCGTCTCGGCTCAAAGCTTAGGA GGACGCAAGCCAGT

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
79	*SOAP6v2*	*UGT74BB1*	*Spinacia oleracea*	Sp_170920_oudh	VIGS	Sp170920_vF	CTAGAGCTCTCGTGGCTCGTATGGGAAAG
80						Sp170920_vR	TTCGAGCTCAGGAACAGATTCAGCCGCAA
81	*SOAP7*	*UGT79K1*	*Spinacia oleracea*	Sp_020820_yeau	VIGS	Sp020820_vF	CTAGAGCTCTAGCCACAAAGCTGGGGATG
82						Sp020820_vR	TTCGAGCTCGGGAGGAATTCAACCTCGGG
83					expression in *N. benthamiana*	Sp020820_GB_1F	GCGCCGTCTCGCTCGAATGGGTAAAACAGTAGCAGCT
84						Sp020820_GB_1R	GCGCCGTCTCGCTCAAAGCTTAATTTGCAGTAAGAAAGCGTTTC
85	*SOAP8*	*UGT79L2*	*Spinacia oleracea*	Sp_113700_suxh	VIGS	Sp113700_vF	CTAGAGCTCGTGATGTTCCCATGGCTTGC
86						Sp113700_vR	TTCGAGCTCAAGGTCGGGCTTATGGGTTG
87					expression in *N. benthamiana*	Sp113700_GB_1F	GCGCCGTCTCGCTCGAATGGGTGGAGAGAAAGAGTTG
88						Sp113700_GB_1R	GCGCCGTCTCGCTCAAAGCCTAAGGAACAAGGGCTTGTAA
89	*SOAP8v2*	*UGT79L1*	*Spinacia oleracea*	Sp_072870_funr	VIGS	Sp072870_vF	CTAGAGCTCGGCTTGCCTTTGGACACTTG
90						Sp072870_vR	TTCGAGCTCAAGACAAGGTCGGGCTTCTG
91	*SOAP9*	*UGT73BS1*	*Spinacia oleracea*	Sp_170320_dmqi	VIGS	Sp170320_vF	CTAGAGCTCGTGTGCTGCTGAGGTTGTTG
92						Sp170320_vR	TTCGAGCTCGTCGCAGTGTACCGGATAGG
93					expression in *N. benthamiana*	Sp170320_GB_1F	GCGCCGTCTCGCTCGAATGGAGCTTTCAAACCCTAGC
94						Sp170320_GB_1R	GCGCCGTCTCGGCGACGAACCACCTTCTTCA
95						Sp170320_GB_2F	GCGCCGTCTCGTCGCGTAACAATTTAAGTGATTTG
96						Sp170320_GB_2R	GCGCCGTCTCGCTCAAAGCTTATTCTGAGTTTGTGGACACTG
97	*BvCSLG*	*CSLG2*	*Beta vulgaris*	XM_010673823.2	VIGS	BvCSL_vF	AAAAGAGCTCCCTTCGGCTCCTCCACAAAA

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
98						BvCSL_vR	AAAAGAGCTCACGAAACTCCGAAAA GGGCT
99					expression in *N. benthamiana*	BvCSL_GB_1F	GCGCCGTCTCGCTCGAATGTCTTCT CTCCACATTTGC
100						BvCSL_GB_1R	GCGCCGTCTCGGCCTCTTTTCCCTG GATACG
101						BvCSL_GB_2F	GCGCCGTCTCGAGGCCAAATCGTC CTCATCG
102						BvCSL_GB_2R	GCGCCGTCTCGCATCTCTTGTCATT TCCTTCAAG
103						BvCSL_GB_3F	GCGCCGTCTCGGATGACTTGTTAGA AGAGGCT
104						BvCSL_GB_3R	GCGCCGTCTCGCTCAAAGCTCAATC ACGTCCTTTTCTTACTTT
105	*CqCSLG*	*CSLG2*	*Chenopodium quinoa*	XM_021866098.1	expression in *N. benthamiana*	CqCSL_GB_1F	GCGCCGTCTCGCTCGAATGGCGGC AACACACATTTG
106						CqCSL_GB_1R	GCGCCGTCTCGGCCTCTTTTCCCTG GCTACA
107						CqCSL_GB_2F	GCGCCGTCTCGAGGCCAGGTCATC CTCATCG
108						CqCSL_GB_2R	GCGCCGTCTCGCTCAAAGCTTATTC TTTCTTTCTAAGTTTGTCTG
109	*MsCSLG*	*CSLG2*	*Medicago sativa*	MSAD_299835	expression in *N. benthamiana*	MsCSL_GB_1F	GCGCCGTCTCGCTCGAATGGCAAC CTTCACATTTCAC
110						MsCSL_GB_1R	GCGCCGTCTCGGCATCTCTCTTGAAATA TTCTGCTTCT
111						MsCSL_GB_2F	GCGCCGTCTCGGGATGTGTAATTTTACA AGAAGCATGTG
112						MsCSL_GB_2R	GCGCCGTCTCGCTCAAAGCCTACC CACTCTTCCGCTTCA
113					silencing in *M. sativa* hairy roots	MsCSL_asGB_1F	GCGCCGTCTCGCTCGAATGCTACC CACTCTTCCGCTTCA
114						MsCSL_asGB_1R	GCGCCGTCTCGCTCAAAGCATGGC AACCTTCACATTTCAC
115	*GmCSL*	*CSLG2*	*Glycine max*	NM_001365113.1	expression in *N. benthamiana*	GmCSL_GB_1F	GCGCCGTCTCGCTCGAATGGCGAC CTTCCACACAGA
116						GmCSL_GB_1R	GCGCCGTCTCGGGATCTCTTCCAAA TTCGTCA

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
117						GmCSL_GB_2F	GCGCCGTCTCGATCCTAAAAATCGT TCCATTGTGT
118						GmCSL_GB_2R	GCGCCGTCTCGCTCAAAGCCTATTG CACCTTGCTTTTCATG
119	*GuCSL*	*CSLG2*	*Glycyrrhiza uralensis*	Glyur003152s0003749 1.1	expression in *N. benthamiana*	GuCSL_GB_1F	GCGCCGTCTCGCTCGAATGGCAAG CTTCACCCTTCA
120						GuCSL_GB_1R	GCGCCGTCTCGATACCAGAGTGTTC ATCACCT
121						GuCSL_GB_2F	GCGCCGTCTCGGTATCTGCCCTTGC CATGGA
122						GuCSL_GB_2R	GCGCCGTCTCGAGATCTCGCTCTGA CCTGAA
123						GuCSL_GB_3F	GCGCCGTCTCGATCTCATCAAGGCT AAATACGAG
124						GuCSL_GB_3R	GCGCCGTCTCGCTCAAAGCCTATCC ACTCTTGCTTTTCATG
125	*GuUGAT*	*UGAT*	*Glycyrrhiza uralensis*	KT759000.1	expression in *N. benthamiana*	GuUGAT_GB_1F	GCGCCGTCTCGCTCGAATGACCAT GGGTAACGAGAAT
126						GuUGAT_GB_1R	GCGCCGTCTCGGCGACGATCCTCC TTCCTCC
127						GuUGAT_GB_2F	GCGCCGTCTCGTCGCACAACGATTT TAACTCTTTA
128						GuUGAT_GB_2R	GCGCCGTCTCGCTCAAAGCTTAATG GGCACGCGACCTCA
129	*GuUGT73P1 2*	*UGT73P12*	*Glycyrrhiza uralensis*	Scaffold00629 (LC314779)	expression in *N. benthamiana*	GuUGT73P12_GB_ 1F	GCGCCGTCTCGCTCGAATGGACTC CTTTGGGGTTGA
130						GuUGT73P12_GB_ 1R	GCGCCGTCTCGGGCCTCATTTCACG GAACAGT
131						GuUGT73P12_GB_ 2F	GCGCCGTCTCGAGGCCAGATTTCAT AGTCACT
132						GuUGT73P12_GB_ 2R	GCGCCGTCTCGCTCAAAGCTTAAGC CACTGCCTCCATTAA
133	**ACYLTRANSFERASES**						
134	*SoAT1*	*uncharacterized acetyltransferase At3g50280-like*	*Spinacia oleracea*	Sp_074630_ygho	VIGS	Spo04549_vF	CTAGAGCTCAGGCGTTGCTATCGAT CCAG

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
135						Spo04549_vR	TTCGAGCTCCATGGCCTTGAGTCTC AGCA
136	*SoAT2*	*O-acyltransferase WSD1-like*	*Spinacia oleracea*	Sp_123780_pgiy	VIGS	Spo21561_vF	CTAGAGCTCCCGGAAATGAAGGTCT GGGT
137						Spo21561_vR	TTCGAGCTCAGGGGGTTCGAGCATT TGTC
138	*SoAT3*	*acyl-CoA-sterol O-acyltransferase 1*	*Spinacia oleracea*	Sp_149180_nwmy	VIGS	Spo15788_vF	CTAGAGCTCTGATAAGGGCCCACTT TGCG
139						Spo15788_vR	TTCGAGCTCCTCCAGTTCGAGCCCT AACAA
140	*SoAT4*	*malonyl-CoA:anthocyanid in 5-O-glucoside-6"-O-malonyltransfera se-like*	*Spinacia oleracea*	Sp_198340_focw	VIGS	Spo13090_vF	CTAGAGCTCAAGCCAAAGCAAAAGG CACC
141						Spo13090_vR	TTCGAGCTCTCGGTTTCCCCCACCA AAAA
142	*SOAP10*	*salutaridinol 7-O-acetyltransferase*	*Spinacia oleracea*	Sp_125800_kzws	VIGS	Spo02253_vF	CTAGAGCTCCAACAACTTCCAAAGG CGGC
143						Spo02253_vR	TTCGAGCTCTCTTGAAGCCCATTGC TGCT
144					expression in *E. coli*	Spo02253_NdeI_F	GAAACATATGGGGGAAGTCAACCAT GAAGAAG
145						Spo02253_NotI_R	CTTTGCGGCCGCCTAATTAGGAGTA GCAAAGCAAGG
146					expression in *N. benthamiana*	Spo02253_GB_1F	GCGCCGTCTCGCTCGAATGGGGGA AGTCAACCATGA
147						Spo02253_GB_1R	GCGCCGTCTCGCTGTCTCAAATTTG GTGGGTT
148						Spo02253_GB_2F	GCGCCGTCTCGACAGTGACAGGGT TTATTTGG

PCT/IL2019/051000

	A	B	C	D	E	F	G
1	**gene name**	**annotation**	**species**	**locus**	**purpose**	**primer name**	**sequence**
149						Spo02253_GB_2R	GCGCCGTCTCGGCCTCCCATCGTT GGGAATT
150						Spo02253_GB_3F	GCGCCGTCTCGAGGCCCCCTAAGG TCAGGAA
151						Spo02253_GB_3R	GCGCCGTCTCGCTCAAAGCCTAATT AGGAGTAGCAAAAGCAAG
152	**OTHERS**						
153	*SoCHLH*	*Magnesium chelatase subunit H*	*Spinacia oleracea*	Sp_082000_pjxp	VIGS	SoCHLH_vF	ACTAGAATTCCAAGTGGGGATGAGT GATGCTTG
154						SoCHLH_vR	GTTCGAATTCCTAGCAGTGCTGATG ATGGAACTC
155	*BvCHLH*	*Magnesium chelatase subunit H*	*Beta vulgaris*	XM_010674548.2	VIGS	BvCHLH_vF	AAAAGAATTCGGAGGCAAGAGGGG CTAAAG
156						BvCHLH_vR	AAAAGAATTCTTTCCTCTTCAAGTCC GCCC
157	*SoUGD1*	*UDP-glucose 6-dehydrogenase 1*	*Spinacia oleracea*	Sp_189830_psca	expression in yeast	SoUGD1_SalI_F	AAAAGTCGACATGGTGAAGAAACTG AAGATTTGCTG
158						SoUGD1_XhoI_R	AAAACTCGAGTTAAGCTACGGCAGG CATGTC

[00927] *Transient expression and candidate gene screening in N. benthamiana*

[00928] pDBG3α and pDBG3Ω constructs were transformed into *Agrobacterium tumafaciens* (GV3101) electrocompetent cells. Transformants were grown on LB plates containing 50 μg/mL kanamycin (pDBG3α) or 200 μg/mL spectinomycin (pDBG3Ω) and 50 μg/mL gentamicin at 28°C. 10 mL of LB medium supplemented with antibiotics was inoculated with a single colony and grew O/N at 28°C. Cells were centrifuged at 3,000 x *g* for 10 min. and supernatant removed. Pellet was resuspended in 5 ml of infiltration medium (100 mM MES buffer, 2mM $Na_3PO_4 \cdot 12H_2O$, 100 μM acetosyringone) and centrifuged again. Pellet was resuspended again in 10 mL of infiltration medium and incubated at room temperature for 2 h. *Agrobacterium* suspensions (OD600 = 0.3 for each strain) were infiltrated into the underside of *N. benthamiana* leaves with a needleless 1 mL syringe. Plants were grown 4-5 weeks under a 16 h light cycle prior to infiltration. Leaves were harvested 4 days post-infiltration, frozen in liquid nitrogen and stored at -80°C for later processing. Biological replicates consisted of several leaves all from different tobacco plants.

[00929] *Expression of SOAP genes in S. cerevisiae WAT 11 and metabolite extraction*

[00930] pESC constructs were transformed into *Saccharomyces cerevisiae* WAT11 using Yeastmaker™ yeast transformation system (Clontech). Yeast cells were transformed with various combinations of pESC vectors allowing expression of one (*SOAP5*), two (*SOAP1-2*), four (*SOAP1-4*), five (*SOAP1-5* or *SOAP1-4+SoUGD1*) and 6 genes (*SOAP1-5+SoUGD1*). Transformed yeast were grown on SD minimal media supplemented with appropriate amino acids and 2% glucose. Colonies were screened and presence of transgene confirmed by colony PCR. For induction of gene expression transformed cells were transferred to minimal medium with 2% galactose and grown for 24h at 30°C. Cultures were centrifuged for 10 min. at 700 x *g*, pellet was resuspended in 1 mL of H_2O, transferred to 2 mL Eppendorf tube and centrifuged again at 8,000 x *g* for a minute. Cell pellet was weighed, equal amount of water and double amount of glass beads (diameter 500 μm) was added and vortexed 5 times for one minute, each time keeping the cells on ice for one minute between vortexings. Lysed cells were mixed with 500 μL of methanol and centrifuged at 14'000 rpm for 5 minutes, clear supernatant was collected and dried in the SpeedVac O/N. Dry residues were dissolved in 150 μL of 80% methanol, filtered through 0.22 μm filter and analyzed on LC-MS.

[00931] *Subcellular localization and confocal microscopy analysis*

[00932] Plasmids allowing expression of studied proteins in fusion with fluorescent protein (FP) were prepared using GoldenBraid system (M. Vazquez-Vilar et al., (2017) *ibid*). SOAP1:GFP, SOAPs 2-4 in fusion with -GFP or -RFP and SOAP5:RFP were used. As cell compartment markers plasmids obtained from ABRC (ER-gk CD3-955 and G-gk CD3-963) were used (B. K. Nelson, X. Cai, A. Nebenführ, A multicolored set of in vivo organelle markers for co-localization studies in

Arabidopsis and other plants. Plant J. 51, 1126-36 (2007)).

[00933] For subcellular localization studies of spinach cellulose synthase like G, SOAP5:RFP was transiently expressed together with ER or Golgi marker in *N. benthamiana* epidermal cells. A bacterial absorbance A600 nm of 0.15 was used for infiltrating each *Agrobacterium* strain. After 72h post infiltration, leaf discs (~0.4 cm diameter) were collected and analyzed for fluorescence with confocal microscopy using the following parameters: Fluorescence was observed by a Nikon eclipse A1 microscope with laser at 488 nm for excitation and images were acquired for GFP (ER or Golgi marker) and 561 nm for red fluorescent protein (RFP) signals. To increase signal-to-noise ratio, each scan pixel was sampled four times and averaged. The same parameters were used for checking colocalization of SOAP genes.

[00934] *Fluorescence Resonance Energy Transfer (FRET) analysis*

[00935] Mean FRET index (i.e., the mean of acceptor, mRFP-fused protein, intensity due to FRET in each pixel after threshold application to remove background noise) were calculated using the FRET Analyzer plugin (M. Hachet-Haas et al., FRET and colocalization analyzer a method to validate measurements of sensitized emission FRET acquired by confocal microscopy and available as an ImageJ Plug-in. Microsc Res Tech. 69, 941-56 (2006)) FIJI/ImageJ. The mean FRET index was calculated for three independent images for each protein combination being tested for proximity/interaction. Cells expressing GFP and RFP only were used to calculate donor and acceptor bleed through.

[00936] *Phylogenetic analysis*

[00937] Homology searches were performed with various query sequences at the non-redundant protein database of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the Blastp option. Protein sequences were aligned using the Muscle algorithm and phylogenetic tree was inferred with RAxML rapid bootstrapping and subsequent ML search (A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312-3 (2014)). Likelihood of final tree was evaluated and optimized under GAMMA model of rate heterogeneity across sites. GAMMA model parameters were estimated up to an accuracy of 0.1000000000 Log Likelihood units. One thousand bootstrap replicates using the fast bootstrap option of RAxML were performed. Used substitution matrix was DAYHOFF. Phylogenetic tree was visualized with iTOL (I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242-5 (2016)).

[00938] *Silencing of MsCSL in M. sativa hairy roots*

[00939] Vector pDBG3α2 with antisense RNA targeting *MsCSL* under 35S promoter and with 35S:KanR cassette was generated using GoldenBraid system. Transformation of *M. sativa* and

generation of transgenic hairy roots was performed as described previously (A. Boisson-Dernier, M. Chabaud, F. Garcia, G. Bécard, C. Rosenberg, D. G. Barker, Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. Mol Plant Microbe Interact. 14, 695-700 (2001)). Instead of Arqua1, ATCC15834 strain of *A. rhizogenes* was used. Transgenic roots were collected, ground and metabolites extracted as described above.

[00940] *Template-based protein structure modeling*

[00941] A homology model of SOAP5 was generated with RaptorX (M. Källberg et al., Template-based protein structure modeling using the RaptorX web server. Nature Protocols. 7, 1511–1522 (2012)) using the crystal structure of *Rhodobacter sphaeroides* cellulose synthase in complex with cyclic-di-GMP and UDP as a template (PDB entry: 4P00) (J. L. Morgan, J. T. McNamara, J. Zimmer, Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. Nat Struct Mol Biol. 21, 489-96 (2014)). Best model was used as a template for MOLE 2.5, a universal toolkit for automated location and characterization of channels, tunnels and pores (L. Pravda et al., MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). Nucleic Acids Res. 46, W368-W373 (2018)). Scheme representing transmembrane topology of SOAP5 was generated with Protter (U. Omasits, C. H. Ahrens, S. Müller, B. Wollscheid, Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics. 30, 884-6 (2014)).

Example 15: Triterpenoid Saponins Biosynthesis Pathway in Spinach

[00942] *Objective:* To discover the triterpenoid saponins biosynthesis pathway in spinach (*Spinacia oleracea*), member of the *Caryophyllales*, particularly enzymes involved in the glycosylation of triterpenoid saponins.

[00943] *Methods:* Liquid chromatography-Mass spectroscopy methods were used in the analysis of saponins in spinach leaves. See Materials and Methods above.

[00944] *Results:*

[00945] Analysis of spinach leaves (using LC-MS) revealed a complex triterpenoid saponin mixture (**FIGURE 18**, **FIGURES 19A-19P**, **Table 12**) comprising more than 20 molecules with medicagenic acid as an aglycone and approximately half of them were acetylated. **Table 13** below provides the details for each compound identified by number in the lower scan of **FIGURE 18**. **Table 12** below uses the following terms: **Sample** - type of experiment/genotype that sample was derived from; **Ret. Time** - Retention time, in minutes; **Putative Name** - putative metabolite identification; **Mol. Formula** - molecular formula of the metabolite or its FA adduct; **Theor. m/z** - theoretical monoisotopic mass calculated for the ion [M-H]-, [M+H]+; **Found m/z** - mass found; **m/z error**

(ppm) - difference between theoretical and found m/z values in ppm; **MS/MS fragments** - fragments, obtained from the ion [M-H]- , [M+H]+; **MS/MS CE (eV)** - collision energy used for fragmentation; **UV/Vis** - UV/Vis absorbance maxima. The UV/Vis spectra (200–600 nm) were acquired on a UPLC (Waters, Acquity) instrument equipped with an Acquity 2996 PDA detector under LC conditions as described in the Materials and Methods. **(S)** - identification confirmed by the standard compound; **(NMR)** - identification confirmed by NMR; **MA** - medicagenic acid; **AA** - augustic acid (2-hydroxy oleanolic acid); **B** - bayogenin; **PG** - polygalagenin; **OA** - oleanolic acid; **G** - gypsogenin; **GA** - gypsogenic acid; **H** - hederagenin; **GhA** - glycyrrhetinic acid; **hexA** - hexuronic acid; **hex** - hexose; **dhex** - deoxyhexose; **pent** - pentose; **GlcA** - glucuronic acid; **Ac** – acetyl.

[00946] **Table 12: Mass Spectroscopy based identification of triterpenoid saponins in studied plants .**

PCT/IL2019/051000

Section A

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Sample	combination of genes expressed in *N. benthamiana*	Rel. Time	Putative Name	Aglycone + sugars	Mol formula	ES(+) Theor m/z $[M+H]^+$	ES(+) Found m/z $[M+H]^+$	m/z error (ppm)	ES(-) Theor m/z $[M-H]^-$	ES(-) Found m/z $[M-H]^-$	m/z error (ppm)	MS/MS ES(-) fragments	MS/MS ES(+) CE (eV)
2	Spinach		-	13.7		MA+dhex+dhex+pent+hexA	C53 H82 O24	-	-	-	1101.5118	1101.5116	-0.2	-	-
3	Spinach		-	12.5	Yossesside VI	MA+dhex+dhex+hexA	C54 H84 O25	-	-	-	1131.5223	1131.5201	-1.9	-	-
4	Spinach		-	19.4		MA+dhex+dhex+Ac+pent+hexA	C55 H84 O25	-	-	-	1143.5223	1143.5285	3.7	-	-
5	Spinach		-	14.7	Yossesside XII	MA+dhex+dhex+pent+pent+hexA	C58 H90 O28	-	-	-	1233.5599	1233.5584	1.1	-	-
6	Spinach		-	12.8	Yossesside IV	MA+dhex+dhex+hex+pent+hexA	C59 H92 O29	-	-	-	1263.5646	1263.5713	0.6	-	-
7	Spinach	WT	-	20.84	Yossesside VII	MA+dhex+dhex+Ac+pent+pent+hexA	C60 H92 O29	-	-	-	1275.5646	1275.5640	-0.5	-	-
8	Spinach	WT	-	15.8	Yossesside Va	MA+dhex+dhex+Ac+hex+pent+hexA	C61 H94 O30	-	-	-	1305.5752	1305.5773	1.6	-	-
9	Spinach	WT	-	18.6	Yossesside V (NMR)	MA+dhex+dhex+Ac+hex+pent+hexA	C61 H94 O30	-	-	-	1305.5752	1305.5781	2.2	-	-
10	Spinach	WT	-	14.2	Yossesside VIII	MA+dhex+dhex+pent+pent+pent+hexA	C63 H98 O32	-	-	-	1365.5963	1365.5981	1.3	-	-
11	Spinach	WT	-	12.4	Yossesside XI	MA+dhex+dhex+hex+pent+pent+hexA	C64 H100 O33	-	-	-	1395.6069	1395.6098	-0.1	-	-
12	Spinach	WT	-	16.9	Yossesside IX	MA+dhex+dhex+Ac+pent+pent+pent+hexA	C65 H100 O33	-	-	-	1407.6069	1407.6107	2.7	-	-
13	Spinach	WT	-	16.7	Yossesside X	MA+dhex+dhex+Ac+hex+pent+pent+hexA	C65 H102 O34	-	-	-	1437.6174	1437.6160	0.6	-	-
14															
15	Spinach		-	16.6		AA+hex+hexA	C42 H66 O15	-	-	-	809.4323	809.4337	1.7	-	-
16	Spinach		-	10.3		B+hex+hexA	C42 H66 O16	-	-	-	825.4273	825.4278	0.6	-	-
17	Spinach		-	15.4		AA+hex+pent+hexA	C47 H74 O19	-	-	-	941.4746	941.4737	-1.0	-	-
18	Spinach		-	15.4		AA+hex+hexA+3-oxopyruvic acid and glycolic acid isomer 1	C47 H70 O21	-	-	-	969.4331	969.4326	-0.5	-	-
19	Spinach		-	16.2		AA+hex+hexA+3-oxopyruvic acid and glycolic acid isomer 2	C47 H70 O21	-	-	-	969.4331	969.4327	-0.4	-	-
20	Spinach		-	13.1		AA+hex+hex+hexA	C48 H78 O20	-	-	-	971.4852	971.4866	1.4	-	-
21	Spinach		-	14.4		AA+hex+pent+pent+hexA	C52 H82 O23	-	-	-	1073.5169	1073.5216	4.4	-	-
22	Spinach		-	11.6		PG+hex+pent+pent+hexA	C52 H80 O24	-	-	-	1087.4961	1087.4990	2.7	-	-
23	Spinach	VIGS -CYPs	-	18.7	Yossesside VIIa	MA+dhex+dhex+Ac+pent+pent+hexA	C60 H92 O29	-	-	-	1275.5646	1275.5675	2.3	-	-
24	Spinach		-	17.2		OA+hex+hexA+3-oxopyruvic acid and glycolic acid (Betavulgaroside I) isomer 1	C47 H70 O20	-	-	-	953.4382	953.4396	-1.7	-	-
25	Spinach		-	18.4		OA+hex+hexA+3-oxopyruvic acid and glycolic acid (Betavulgaroside I) isomer 2	C47 H70 O20	-	-	-	953.4382	953.4369	-1.4	-	-
26	Spinach		-	14.6		G+hex+hexA-3-oxopyruvic acid and glycolic acid (Basellasaponin B)	C47 H68 O21	-	-	-	967.4175	967.4183	0.8	-	-
27	Spinach		-	16.1		GA+hex+pent+dhex+dhex+hexA	C59 H92 O28	-	-	-	1247.5697	1247.5695	-0.2	-	-
28	Spinach		-	21.2		GA+hex+pent+dhex+dhex+Ac+hexA	C61 H94 O29	-	-	-	1289.5803	1289.5781	-1.7	-	-
29															
30	Spinach		-	10.74		MA+hex+hexA+3-oxopyruvic acid and glycolic acid isomer 1	C47 H68 O23	-	-	-	999.4073	999.4097	2.4	-	-
31	Spinach		-	11.90		MA+hex+hexA+3-oxopyruvic acid and glycolic acid isomer 2	C47 H68 O23	-	-	-	999.4081	999.4097	1.6	-	-
32	Spinach		-	15.30		MA+dhex+hexA+3-oxopyruvic acid and glycolic acid	C47 H68 O22	-	-	-	983.4124	983.4155	3.2	-	-
33	Spinach		-	15.31		MA+hexA+dhex+pent (Betavulgaroside III)	C47 H72 O20	-	-	-	955.4539	955.4543	0.4	-	-
34	Spinach		-	11.00		MA+pent+hex+hexA	C47 H72 O21	-	-	-	971.4488	971.4506	1.9	-	-
35	Spinach		-	13.40		MA+dhex+dhex+hexA	C48 H74 O20	-	-	-	969.4695	969.4661	-3.5	-	-
36	Spinach		-	17.10		OA+hexA+hex+pent	C47 H74 O18	-	-	-	925.4797	925.4789	-0.9	-	-
37	Spinach		-	12.50		AA/H+hexA+hex	C42 H66 O15	-	-	-	809.4323	809.4335	1.5	-	-
38	Spinach	VIGS - GTs +CSL	-	14.62		G+hexA+hex	C42 H64 O15	-	-	-	807.4167	807.4156	-1.4	-	-
39	Spinach		-	12.90		**GA**/PG+hex+hexA	C42 H64 O16	-	-	-	823.4116	823.4129	1.6	-	-
40	Spinach		-	10.00		MA+dhex+pent+pent+hex+hexA	C58 H90 O29	-	-	-	1245.5490	1249.5537	3.8	-	-
41	Spinach		-	8.50		MA+dhex+hex+hex+hexA	C54 H84 O26	-	-	-	1147.5173	1147.5226	4.6	-	-
42	Spinach		-	9.80		MA+dhex+pent+hex+hexA	C53 H82 O25	-	-	-	1117.5067	1117.5078	1.0	-	-
43	Spinach		-	11.00		MA+hex+hexA	C42 H64 O17	-	-	-	839.4065	839.4077	1.4	-	-
44	Spinach		-	21.10		MA+pent+hexA	C41 H62 O16	-	-	-	809.3960	809.3938	-2.7	-	-
45	Spinach		-	15.25		**PG**/GA+hex+hexA	C42 H64 O16	-	-	-	823.4116	823.4132	1.9	-	-
46	Spinach		-	13.20		MA+dhex+dhex+pent+pent+hexA	C58 H90 O28	-	-	-	1233.5540	1233.5499	3.3	-	-
47	Spinach		-	17.60		MA+dhex+dhex+Ac+pent+pent+hexA	C60 H92 O29	-	-	-	1275.5646	1275.5627	-1.5	-	-
48	Spinach		-	11.15		MA+hex+hex	C42 H66 O16	-	-	-	825.4273	825.4283	1.2	-	-
49	Spinach		-	22.73	Medicagenic acid (S)	MA	C30 H46 O6	503.3373	503.3365	-1.6	501.3216	501.3217	0.2	see Fig. S19	25
50															
51	Medicago sativa		-	21.61	Saponin 1	B+hex+hex+hexA	C46 H76 O21	-	-	-	987.4801	987.4788	-1.3	-	-
52	Medicago sativa		-	25.63	Saponin 2	B+hex+hexA	C42 H66 O16	-	-	-	825.4273	825.4273	0.0	-	-
53	Medicago sativa	Differential signals form *M. sativa*	-	23.04	Saponin 3	Soyasapogenol A+dhex+hex+hexA	C48 H78 O19	-	-	-	957.5059	957.5059	-0.2	-	-
54	Medicago sativa		-	24.11	Saponin 4	GA/PG+hex+hexA-malonyl	C45 H66 O19	-	-	-	909.4120	909.4120	0.0	-	-

PCT/IL2019/051000

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Sample	combination of genes expressed in *N. benthamiana*	Rel. Time	Putative Name	Aglycone + sugars	Mol formula	ES(+) Theor m/z [M+H]+	ES(+) Found m/z [M+H]+	m/z error (ppm)	ES(-) Theor m/z [M-H]-	ES(-) Found m/z [M-H]-	m/z error (ppm)	MS/MS ES(-) fragments	MS/MS ES(+) CE (eV)
55	Medicago sativa	hairy roots with silenced *MsCSL*	-	25.23	Saponin 5	B+2hex+hex+HexA	C48 H76 O20	-	-	-	971.4852	971.4849	-0.3	-	-
56	Medicago sativa		-	22.84	Saponin 6	H+Hex+Hex+dHex+HexA	C54 H86 O24	-	-	-	1117.5431	1117.5420	-1.0	-	-
57	Medicago sativa		-	22.92	Saponin 7	GA/PG+hex+HexA	C42 H64 O16	-	-	-	823.4116	823.4078	-4.6	-	-
58	Medicago sativa		-	28.14	Saponin 8	Soyasapogenol B+2hex+Hex+Hex+HexA	C54 H88 O23	-	-	-	1103.5638	1103.5635	-0.3	-	-
59															
60	Beta vulgaris	Differential signals from *B. vulgaris* with silenced *BvCSL*		14.95	Bv-Saponin 1 (beta-vulgarisside V)	OA+HexA+hex+dHex+Act	C53 H82 O25	-	-	-	1117.5060	1117.5060	-0.6	-	-
61	Beta vulgaris			15.24	Bv-Saponin 2	OA+HexA+hex+dHex+pent	C53 H84 O23	-	-	-	1087.5325	1087.5316	-0.8	-	-
62	Beta vulgaris			15.67	Bv-Saponin 3 (beta-vulgarisside IX)	OA+HexA+hex+pent+Act	C52 H80 O24	-	-	-	1087.4961	1087.4978	1.6	-	-
63	Beta vulgaris			16.95	Bv-Saponin 4	OA+hex+pent+hexA	C47 H74 O18	-	-	-	925.4772	925.4782	-1.6	-	-
64	Beta vulgaris			17.54	Bv-Saponin 5 (beta-vulgarisside III)	OA+HexA+hex+Act	C47 H72 O20	-	-	-	955.4539	955.4536	-0.3	-	-
65															
66	Expression in *N. benthamiana*	genes from spinach	SOAP1-4+SOAP5	20.98	medicagenic acid 3-*O*-glucuronide (NMR)	MA+GlcA	C36 H54 O12	-	-	-	677.3543	677.3522	-2.2	-	-
67	Expression in *N. benthamiana*		SOAP1-5+SOAP6	14.96	Yossesicis I	MA+GlcA+Fuc	C42 H64 O16	-	-	-	823.4116	823.4104	-1.5	-	-
68	Expression in *N. benthamiana*		SOAP1-6+SOAP7	13.21	Yossesicis II	MA+GlcA+Fuc+Rha	C48 H74 O20	-	-	-	969.4693	969.4693	-0.2	-	-
69	Expression in *N. benthamiana*		SOAP1-7+SOAP8	12.08	Yossesicis III	MA+GlcA+Fuc+Rha+Glc	C54 H84 O25	-	-	-	1131.5200	1131.5199	-0.1	-	-
70	Expression in *N. benthamiana*		SOAP1-8+SOAP9	12.37	Yossesicis IV	MA+GlcA+Fuc+Rha+Glc+Xyl	C59 H92 O29	-	-	-	1263.5646	1263.5649	0.2	-	-
71	Expression in *N. benthamiana*		SOAP1-9+SOAP10	18.18	Yossesicis V	MA+GlcA+Fuc+Ac+Rha+Glc+Xyl	C61 H94 O30	-	-	-	1305.5752	1305.5736	-1.2	-	-
72															
73	Expression in *N. benthamiana*	genes from *G. uralensis*	bAS+CYP72A154+CYP88D6+GuCSL	17.2	glycyrrhetinic acid 3-*O*-glucuronide	GhA+GlcA	C36 H54 O10	-	-	-	645.3644	645.3644	0.1	-	-
74	Expression in *N. benthamiana*		bAS+CYP72A154+CYP88D6+GuCSL	19.0	glycyrrhetinic acid 3-*O*-glucuronide	GhA+GlcA	C36 H54 O10	-	-	-	645.3644	645.3636	1.2	-	-
75															
76	Expression in *N. benthamiana*		bAS+CYP72A154+CYP88D6+SOAP5	17.2	glycyrrhetinic acid 3-*O*-glucuronide	GhA+GlcA	C36 H54 O10	-	-	-	645.3644	645.3648	-0.6	-	-
77	Expression in *N. benthamiana*		bAS+CYP72A154+CYP88D6+SOAP5	19.0	glycyrrhetinic acid 3-*O*-glucuronide	GhA+GlcA	C36 H54 O10	-	-	-	645.3644	645.3639	-2.3	-	-
78															
79	Expression in *N. benthamiana*		bAS+CYP72A154+CYP88D6+GuCSL+UGT73P12	11.8	glycyrrhizin isomer 1 (S)	GhA+GlcA+GlcA	C42 H62 O16	-	-	-	821.3965	821.3963	0.3	-	-
80	Expression in *N. benthamiana*		bAS+CYP72A154+CYP88D6+GuCSL+UGT73P12	13.4	glycyrrhizin isomer 2 (S)	GhA+GlcA+GlcA	C42 H62 O16	-	-	-	821.3965	821.3960	0.6	-	-
81															
82	Expression in *N. benthamiana*		SOAP1-4+GuCSL	14.2	medicagenic acid 3-*O*-glucuronide (S)	MA+GlcA	C36 H54 O12	-	-	-	677.3543	677.3535	1.0	-	-

PCT/IL2019/051000

Section B

	A	B	P	Q	R	S	T
1		Sample	MS/MS ES(-) fragments	MS/MS ES(-) CE (eV)	UV/Vis	Detected in other species	References
2	Spinach		969.4370 [M-pent-H]-; 955.4099 [M-dhex-H]-; 793.4330 [M-hexA-pent-H]-; 501.3204 [M-hexA-pent-dhex-dhex-H]-				
3	Spinach		955.4596 [M-hexA-H]-; 823.4120 [M-hex-dhex-H]-; 677.3536 [M-hex-dhex-dhex-H]-; 501.3264 [M-hex-dhex-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
4	Spinach		1011.4806 [M-pent-H]-; 835.4438 [M-pent-HexA-H]-;	ramp 15-50 eV	-	-	-
5	Spinach		926.4760 [M-hexA-pent-H]-; 501.3236 [M-hexA-pent-dhex-dhex-H]-	ramp 15-50 eV	70	-	-
6	Spinach		955.4892 [M-hexA-pent-dhex-H]-; 809.3924 [M-hexA-pent-hex-dhex-dhex-H]-		70	-	-
7	Spinach	WT	1143.5160 [M-pent-]-; 1011.4427 [M-pent-pent-H]-; 967.4788 [M-pent-hex-hexA-H]-; 823.4001 [M-pent-pent-dhex-dhex-Ac-dhex-H]-; 501.3212 [M-pent-pent-dhex-dhex-hexA-H]-	ramp 15-50 eV	70	-	-
8	Spinach		997.4596 [M-hexA-pent-H]-; 955.4602 [M-hexA-pent-Ac-H]-; 501.3232 [M-hexA-pent-dhex-dhex-Ac-hex-H]-		70	-	-
9	Spinach		997.4596 [M-hexA-pent-Ac-H]-; 955.4875 [M-hexA-pent-dhex-dhex-Ac-hex-H]-; 501.3210 [M-hexA-pent-dhex-dhex-Ac-hex-H]-		70	-	-
10	Spinach		1057.5228 [M-hexA-pent-H]-; 926.4678 [M-hexA-pent-pent-H]-; 647.3792 [M-hexA-pent-pent-pent-dhex-dhex-H]-; 501.3193 [M-hexA-pent-pent-pent-dhex-dhex-H]-		70	-	-
11	Spinach		1263.5773 [M-pent-H]-; 1087.5323 [M-hexA-pent-H]-; 955.4877 [M-hexA-pent-pent-H]-; 809.3959 [M-hexA-pent-pent-dhex-dhex-hex-H]-; 501.3225 [M-hexA-pent-pent-dhex-dhex-hex-H]-		70	-	-
12	Spinach		1099.5310 [M-hexA-pent-Ac-H]-; 1057.5170 [M-hexA-pent-Ac-H]-; 501.3190 [M-hexA-pent-pent-dhex-Ac-dhex-H]-		70	-	-
13	Spinach		1129.5420 [M-pent-hexA-Ac-H]-; 1087.5226 [M-pent-hexA-pent-hex-dhex-Ac-dhex-H]-		70	-	-
14							
15	Spinach		647.3809 [M-hex-H]-; 603.3903 [M-hex-CO2-H2C-H]-; 471.3475 [M-hex-hexA-H]-		60	-	-
16	Spinach		663.3741 [M-hex-H]-; 619.3843 [M-hex-CO2-H]-; 601.3750 [M-hex-CO2-H2C-H]-; 487.3431 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
17	Spinach		809.4344 [M-oxopyruvic and glycolic acid-H]-; 647.3788 [M-hex-oxopyruvic and glycolic acid-H]-; 585.3795 [M-hex-oxopyruvic and glycolic acid-CO2-H2O-H]-; 471.3475 [M-hex-oxopyruvic and glycolic acid-hexA-H]-	ramp 15-50 eV	-	-	-
18	Spinach		809.4325 [M-oxopyruvic and glycolic acid-H]-; 647.3788 [M-hex-oxopyruvic and glycolic acid-H]-; 585.3802 [M-hex-oxopyruvic and glycolic acid-CO2-H2O-H]-; 471.3477 [M-hex-oxopyruvic and glycolic acid-hexA-H]-	ramp 15-50 eV	60	-	-
19	Spinach		809.4267 [M-hex-H]-; 417.3479 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
20	Spinach		911.4585 [M-hex-H]-; 647.3829 [M-hex-pent-pent-H]-; 471.3446 [M-hex-pent-pent-hexA-H]-	ramp 15-50 eV	-	-	-
21	Spinach		955.4252 [M-pent-H]-; 925.4381 [M-hexA-H]-; 617.3692 [M-hex-pent-hexA-H]-; 485.3267 [M-hex-pent-pent-hexA-H]-	ramp 15-50 eV	-	-	-
22	Spinach	VIGS -CYPs	1143.5077 [M-pent-]-; 1011.4468 [M-pent-pent-H]-; 677.3535 [M-pent-pent-dhex-Ac-dhex-hexA-H]-; 501.3269 [M-pent-pent-dhex-Ac-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
23	Spinach		793.4354 [M-oxopyruvic and glycolic acid-H]-; 631.3831 [M-oxopyruvic and glycolic acid-hex-H]-; 569.3835 [M-oxopyruvic and glycolic acid-hex-CO2-H2O-H]-; 455.3514 [M-oxopyruvic and glycolic acid-hex-hexA-H]-		60	-	*Beta vulgaris / Basella rubra* (54)
24	Spinach		793.4374 [M-oxopyruvic and glycolic acid-H]-; 631.3844 [M-oxopyruvic and glycolic acid-hex-H]-; 569.3845 [M-oxopyruvic and glycolic acid-hex-CO2-H2O-H]-; 455.3517 [M-oxopyruvic and glycolic acid-hex-hexA-H]-	ramp 15-50 eV	-	*Beta vulgaris / Basella rubra*	(54)
25	Spinach		807.4161 [M-oxopyruvic and glycolic acid-H]-; 645.3607 [M-oxopyruvic and glycolic acid-hex-H]-; 583.3608 [M-oxopyruvic and glycolic acid-hex-CO2-H2O-H]-; 469.3307 [M-oxopyruvic and glycolic acid-hex-hexA-H]-		60	-	*Basella rubra* (55)
26	Spinach		1115.5173 [M-pent-H]-; 969.4326 [M-dhex-pent-H]-; 939.4197 [M-hex-dhex-H]-; 823.4352 [M-dhex-pent-dhex-H]-; 485.3267 [M-pent-hex-dhex-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
27	Spinach		969.4322 [M-dhex-pent-Ac-H]-; 939.4189 [M-hex-dhex-Ac-H]-; 823.4349 [M-dhex-pent-dhex-Ac-H]-; 485.3265 [M-pent-hex-dhex-dhex-hexA-Ac-H]-	ramp 15-50 eV	-	-	-
28	Spinach						
29							
30	Spinach		955.4204 [M-CO2-H]-; 897.4148 [M-CO2-C2H2O2-H]-; 839.4088 [M-CO2-C2H2O2-C2H2O2-H]-; 677.3586 [M-oxopyruvic and glycolic acid-hex-H]-; 663.3751 [M-oxopyruvic and glycolic acid-hexA-H]-; 501.3274 [M-oxopyruvic and glycolic acid-hex-hexA-H]-	ramp 15-50 eV	-	-	-
31	Spinach		955.4201 [M-CO2-H]-; 897.4130 [M-CO2-C2H2O2-H]-; 839.4070 [M-CO2-C2H2O2-C2H2O2-H]-; 677.3503 [M-oxopyruvic and glycolic acid-hex-H]-; 663.3727 [M-oxopyruvic and glycolic acid-hexA-H]-; 501.3252 [M-oxopyruvic and glycolic acid-hex-hexA-H]-	ramp 15-50 eV	-	-	-
32	Spinach		939.4240 [M-CO2-H]-; 881.4178 [M-CO2-C2H2O2-H]-; 823.4127 [M-CO2-C2H2O2-C2H2O2-H]-; 677.3565 [M-oxopyruvic and glycolic acid-dhex-H]-; 647.3713 [M-oxopyruvic and glycolic acid-hexA-H]-; 501.3176 [M-oxopyruvic and glycolic acid-hex-hexA-H]-	ramp 15-50 eV	-	-	-
33	Spinach		647.3743 [M-hexA-pent-H]-; 501.3240 [M-hexA-pent-dhex-H]-		50	*Beta vulgaris*	(56)
34	Spinach		663.3754 [M-hexA-pent-H]-; 501.3222 [M-hexA-pent-hex]		50	-	-
35	Spinach		823.4380 [M-dhex-H]-; 677.3533 [M-dhex-dhex-H]-; 501.3225 [M-dhex-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
36	Spinach		793.4344 [M-pent-H]-; 631.3828 [M-pent-hex-H]-; 569.3834 [M-pent-hex-CO2-H2O-H]-; 455.3513 [M-pent-hex-hexA-H]-	ramp 15-50 eV	-	-	-
37	Spinach		647.3774 [M-hex-H]-; 471.3475 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
38	Spinach	VIGS - GTs +CSL	645.3638 [M-hex-H]-; 469.3323 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
39	Spinach		661.3693 [M-hex-H]-; 647.3861 [M-hexA-H]-; 617.3719 [M-hex-CO2-H]-; 599.3574 [M-hex-CO2-H2O-H]-; 485.3271 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
40	Spinach		1117.5071 [M-pent-H]-; 985.4347 [M-pent-pent-H]-; 809.4266 [M-pent-hex-dhex-H]-; 501.3194 [M-hexA-hex-pent-pent-dhex-H]-	ramp 15-50 eV	-	-	-
41	Spinach		985.4608 [M-hex-H]-; 823.3881 [M-hex-hex-H]-; 677.3605 [M-hex-hex-dhex-H]-; 501.3212 [M-hex-hex-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
42	Spinach		985.4552 [M-pent-H]-; 955.4471 [M-hex-H]-; 677.3557 [M-pent-hex-dhex-H]-; 501.3250 [M-pent-hex-dhex-hexA-H]-	ramp 15-50 eV	-	-	-
43	Spinach		677.3486 [M-hex-H]-; 501.3224 [M-hex-hexA-H]-	ramp 15-50 eV	-	-	-
44	Spinach		501.3207 [M-hexA-pent-H]-; 325.0773 [hexA-pent-H2O-H]-	ramp 15-50 eV	-	-	-
45	Spinach		661.3582 [M-hex-H]-; 647.3805 [M-hexA-H]-; 617.3690 [M-hex-CO2-H]-; 599.3580 [M-hex-CO2-H2O-H]-; 485.3252 [M-hex-hexA-H]-		50	-	-
46	Spinach		926.4778 [M-hexA-pent-H]-; 501.3233 [M-hexA-pent-dhex-dhex-pent-H]-		70	-	-
47	Spinach		1143.5116 [M-pent-H]-; 1101.5002 [M-pent-Ac-H]-; 955.4487 [M-pent-Ac-dhex-H]-; 967.4825 [M-pent-hexA-pent-hexA-pent-Ac-H]-; 501.3212 [M-dhex-dhex-Ac-pent-pent-hexA-H]-	ramp 15-50 eV	-	-	-
48	Spinach		663.3739 [M-hex-H]-; 501.3222 [M-hex-hex-H]-	ramp 15-50 eV	-	-	-
49	Spinach				-	-	-
50							
51	Medicago sativa		825.4302 [M-hex-H]-; 663.3773 [M-hex-hex-H]-; 487.3424 [M-hex-hex-hexA-H]-		45	*Medicago truncatula*	*(Main text reference 14)*
52	Medicago sativa		663.3765 [M-hex-H]-; 487.3432 [M-hex-hexA-H]-		45	*Medicago truncatula*	*(Main text reference 14)*
53	Medicago sativa	Differential signals form M. sativa	939.4989 [M-H2O-H]-; 895.5098 [M-CO2-H2O-H]-; 811.4492 [M-dhex-H]-; 767.4595 [M-dhex-CO2-H]-; 749.4526 [M-dhex-CO2-H2O-H]-; 631.3887 [M-dhex-hex-H2O-H]-; 613.3731 [M-dhex-hex-H2O-H2O-H]-; 473.3685 [M-hex-dhex-hex-hexA-H]-		45	*Medicago truncatula*	*(Main text reference 14)*
54	Medicago sativa		865.4246 [M-CO2-H]-; 823.4179 [M-malonyl-H]-; 805.3989 [M-malonyl-H2O-H]-; 703.3777 [M-hex-CO2-H]-; 643.3525 [M-malonyl-hex-H2O-H]-; 599.3606 [M-malonyl-hex-CO2-H2O-H]-; 485.3273 [M-malonyl-hex-hexA-H]-		45	*Medicago truncatula*	*(Main text reference 14)*

PCT/IL2019/051000

	A	B	P	Q	R	S	T
1		Sample	MS/MS ES(-) fragments	MS/MS ES(-) CE (eV)	UV/Vis	Detected in other species	References
55	Medicago sativa	hairy roots with silenced *MsCSL*	953.4714 [M-H2O-H]-; 909.4915 [M-CO2-H2O-H]-; 825.4288 [M-dhex-H]-; 763.4302 [M-dhex-CO2-H2O-H]-; 645.3668 [M-dhex-hex-H2O-H]-; 601.3768 [M-dhex-hex-H2O-CO2-H]-; 555.3726 [M-dhex-hex-106-H]-; 457.3241 [M-dhex-hex-hex-H]-;	60	-	*Medicago truncatula*	*(Main text reference: 14)*
56	Medicago sativa		1099.5315 [M-H2O-H]-; 955.4852 [M-hex-H]-; 893.5015 [M-hex-CO2-H2O-H]-; 791.4280 [M-hex-dhex-H2O-H]-; 747.4368 [M-hex-dhex-CO2-H2O-H]-; 729.4260 [M-hex-dhex-H2O-H2O-CO2-H]-; 629.3726 [M-hex-dhex-hex-H]-; 585.3829 [M-hex-dhex-hex-CO2-H]-; 539.3792 [M-hex-dhex-hex-106-H]-; 471.3504 [M-hex-dhex-hex-hexA-H]-;	60	-	*Medicago truncatula*	*(Main text reference: 14)*
57	Medicago sativa		661.3595 [M-hex-H]-; 617.3698 [M-hex-CO2-H]-; 599.3587 [M-hex-CO2-H2O-H]-; 485.3268 [M-hex-hexA-H]-;	45	-	*Medicago truncatula*	*(Main text reference: 14)*
58	Medicago sativa		1085.5556 [M-H2O-H]-; 1041.5636 [M-H2O-CO2-H]-; 957.5061 [M-dhex-H]-; 895.5071 [M-hex-H2O-CO2-H]-; 795.4548 [M-dhex-hex-H]-; 777.4427 [M-dhex-hex-H2O-H]-; 759.4297 [M-dhex-hex-H2O-H2O-H]-; 457.3681 [M-hex-dhex-dhex-hexA-H]-	45	-	*Medicago truncatula*	*(Main text reference: 14)*
59							
60	Beta vulgaris	Differential signals from *B. vulgaris* with silenced *BvCSL*	997.8342 [M-C9H4O5-H]-; 955.8525 [M-Act-H]-; 835.4744 [M-C9H4O5-hex-H]-; 793.4632 [M-Act-hex-H]-; 731.4594 [M-Act-hex-CO2-H2O-H]-;	60	-	*Beta vulgaris*	*(57)*
61	Beta vulgaris		633.4028 [M-Act-hex-hex-H]-; 613.3919 [M-Act-hex-hex-H2O-H]-; 455.3650 [M-Act-hex-hex-hexA-H]-; 955.5191 [M-pent-H]-; 925.5096 [M-hex-H]-; 793.4595 [M-hex-CO2-pent-H2O-H]-; 613.3902 [M-hex-pent-hex-H2O-H]-;	60	-	*Beta vulgaris*	*(57)*
62	Beta vulgaris		569.4006 [M-hex-pent-hex-H2O-CO2-H]-; 551.3894 [M-hex-pent-hex-H2O-CO2-H2O-H]-; 455.3650 [M-hex-pent-hex-hexA-H]-; 967.5213 [M-C9H4O5-H]-; 925.5092 [M-Act-H]-; 805.4618 [M-C9H4O5-hex-H]-; 763.4507 [M-hex-Act-H]-; 743.4590 [M-C9H4O5-hex-H2O-CO2-H]-;	60	-	*Beta vulgaris*	*(57)*
63	Beta vulgaris		593.3996 [M-C9H4O5-hex-H2O-CO2-pent-H2O-H]-; 455.3639 [M-Act-hex-pent-hexA-H]-; 763.4516 [M-hex-H]-; 701.4487 [M-hex-CO2-H2O-H]-; 631.4028 [M-hex-pent-H]-; 569.4016 [M-hex-CO2-H2O-pent-H]-; 551.3904 [M-hex-CO2-H2O-pent-H2O-H]-; 455.3654 [M-hex-pent-hexA-H]-	60	-	*Beta vulgaris*	*(57)*
64	Beta vulgaris		835.4753 [M-C9H4O5-H]-; 793.4630 [M-Act-H]-; 673.4167 [M-C9H4O5-hex-H]-; 631.4045 [M-Act-hex-H]-; 569.4015 [M-Act-hex-CO2-H2O-H]-; 455.3655 [M-hex-hexA-Act-H]-	60	-	*Beta vulgaris*	*(57)*
65							
66	Expression in N. benthamiana	genes from spinach	501.3206 [M-GlcA-H]-; 483.3096 [M-GlcA-H2O-H]-; 439.3184 [M-GlcA-H2O-CO2-H]-; 193.0346 [GlcA-H]-; 175.0238 [GlcA-H2O-H]-	45	-	-	-
67	Expression in N. benthamiana		647.3776 [M-hexA-H]-; 501.3237 [M-hexA-dhex-H]-	45	-	-	-
68	Expression in N. benthamiana		793.4233 [M-hexA-H]-; 501.3219 [M-hexA-dhex-dhex-H]-; 439.3192 [M-hexA-dhex-dhex-CO2-H2O-H]-	50	-	-	-
69	Expression in N. benthamiana		955.4902 [M-hexA-H]-; 677.3536 [M-hexA-dhex-dhex-dhex-H]-; 501.3195 [M-hexA-dhex-dhex-dhex-H]-	55	-	-	-
70	Expression in N. benthamiana		955.4648 [M-hexA-pent-H]-; 793.4109 [M-hexA-pent-hex-H]-; 501.3018 [M-hexA-pent-hex-dhex-dhex-H]-; 483.3283 [M-hexA-pent-hex-dhex-dhex-H2O-H]-	70	-	-	-
71	Expression in N. benthamiana		997.4995 [M-hexA-pent-H]-; 955.4883 [M-hexA-pent-Ac-H]-; 501.3227 [M-hexA-pent-hex-dhex-dhex-Ac-H]-; 439.3179 [M-hexA-pent-hex-dhex-dhex-Ac-CO2-H2O-H]-	70	-	-	-
72							
73	Expression in N. benthamiana	genes from *G. uralensis*	583.3629 [M-CO2-H2O-H]-; 569.3469 [M-C2O3H4-H; ^{0,5}X]-; 523.3415 [M-C3O4H6-H2O-H; ^{0,3}X-H2O]-; 469.3315 [M-GlcA-H]-; 425.3419 [M-GlcA-CO2-H]-	55	-	*Glycyrrhiza uralensis*	*(58)*
74	Expression in N. benthamiana		583.3636 [M-CO2-H2O-H]-; 569.3470 [M-C2O3H4-H; ^{0,5}X]-; 523.3423 [M-C3O4H6-H2O-H; ^{0,3}X-H2O]-; 469.3318 [M-GlcA-H]-; 425.3419 [M-GlcA-CO2-H]-	55	-	*Glycyrrhiza uralensis*	*(58)*
75							
76	Expression in N. benthamiana		583.3631 [M-CO2-H2O-H]-; 569.3468 [M-C2O3H4-H; ^{0,5}X]-; 523.3418 [M-C3O4H6-H2O-H; ^{0,3}X-H2O]-; 469.3317 [M-GlcA-H]-; 425.3422 [M-GlcA-CO2-H]-	55	-	*Glycyrrhiza uralensis*	*(58)*
77	Expression in N. benthamiana		583.3633 [M-CO2-H2O-H]-; 569.3470 [M-C2O3H4-H; ^{0,5}X]-; 523.3421 [M-C3O4H6-H2O-H; ^{0,3}X-H2O]-; 469.3320 [M-GlcA-H]-; 425.3415 [M-GlcA-CO2-H]-	55	-	*Glycyrrhiza uralensis*	*(58)*
78							
79	Expression in N. benthamiana		645.3636 [M-GlcA-H]-; 469.3303 [M-GlcA-GlcA-H]-; 351.0574 [GlcA-GlcA-H]-; 193.0357 [GlcA-H]-; 175.0258 [GlcA-H2O-H]-	85	-	*Glycyrrhiza uralensis*	*(58)*
80	Expression in N. benthamiana		645.3615 [M-GlcA-H]-; 469.3323 [M-GlcA-GlcA-H]-; 351.0577 [GlcA-GlcA-H]-; 193.0354 [GlcA-H]-; 175.0250 [GlcA-H2O-H]-	85	-	*Glycyrrhiza uralensis*	*(58)*
81							
82	Expression in N. benthamiana		501.3212 [M-GlcA-H]-; 483.3084 [M-GlcA-H2O-H]-; 439.3176 [M-GlcA-H2O-CO2-H]-; 193.0334 [GlcA-H]-; 175.0242 [GlcA-H2O-H]-	45	-	-	-

[00947] **Table 13: Summary of Saponin Composition in Spinach**

No.	Name	Molecular Formula	*m/z* measured $[M-H]^-$	*m/z* calculated $[M-H]^-$	$\Delta m/z$ ppm
1	Yossoside XI	C64H100O33	1395.6068	1395.6089	-0.1
2	Yossoside VI	C47H88O30	1131.5261	1131.5282	-1.9
3	Yossoside IV	C52H96O34	1263.5713	1263.5705	0.6
4	Yossoside VIII	C63H98O32	1365.5981	1365.5963	1.3
5	Yossoside XII	C58H90O28	1233.5554	1233.5540	1.1
6	Yossoside Va	C61H94O30	1305.5773	1305.5752	1.6
7	Yossoside X	C66H102O34	1437.6160	1437.6151	0.6
8	Yossoside VII	C60H92O29	1275.5675	1275.5646	2.3
9	Yossoside V	C61H94O30	1305.5781	1305.5752	2.2
10	Yossoside IX	C65H100O33	1407.6107	1407.6089	2.7
11	Yossoside VIIa	C60H92O29	1275.5640	1275.5646	-0.5

[00948] Mass spectrometry fragmentation analysis and a series of 1D and 2D-NMR experiments established the structure of the most abundant spinach saponin (termed here Yossoside V). It comprises medicagenic acid (the aglycone) with glucuronic acid and xylose attached to the hydroxyl at the C-3 position and acetyl-fucose, rhamnose and glucose linked to the carboxyl at C-28 (**FIGURES 20A-20C**, **FIGURE 21**).

[00949] *Summary:* Eleven triterpenoid saponins were identified in spinach, wherein Yossiside V was the most abundant.

Example 16: Identification of Saponin Biosynthesis Genes Using Spinach Transcriptome Data

[00950] *Objective*: To identify triterpenoid aglycone biosynthesis genes in the triterpenoid saponin biosynthetic pathway.

[00951] *Methods*: See Materials and Methods above.

[00952] *Results*: Transcriptome data was generated from five spinach tissues exhibiting varying content of saponins, and gene candidates selected based on their differential expression between samples with high and low saponin content, as well as their homology to known triterpenoid aglycone biosynthesis genes (E. Biazzi *et al.*, CYP72A67 Catalyzes a Key Oxidative Step in *Medicago truncatula* Hemolytic Saponin Biosynthesis. *Molecular Plant*. **8**, 1493–1506 (2015)).

[00953] Inspecting the genomic location of three of the candidates, including: saponin β-amyrin synthase (*SobAS*; termed SOAP1; Sp_107620_kpnh; SEQ ID NO: 45 [gene sequence] SEQ ID NO: 48 [polypeptide sequence]), cytochrome P450 CYP716A268 (SOAP2; Sp_107660_kiqg; SEQ ID NO: 46 [gene sequence], SEQ ID NO: 49 [polypeptide sequence]), and CYP716A268v2 (SOAP2-like; Sp_107670_ptqx; SEQ ID NO: 47 [gene sequence] SEQ ID NO: 6 [polypeptide sequence])

revealed their close physical genomic location in the form of a metabolic gene cluster (**FIGURES 22A and 22B**).

[00954] Virus Induced Gene Silencing (VIGS) assays of SOAP1 and SOAP2 in spinach resulted in reduced saponin content (**FIGURES 23A-23D**), while their over-expression in *N. benthamiana* resulted the production of β-amyrin and oleanolic acid (**FIGURE 24A and 24B**).

[00955] *Summary*: These results provided strong evidence for the involvement of these genes in spinach saponin biosynthesis.

Example 17: Co-Expression Analysis for Identification of Additional Saponin Biosynthesis Genes & Characterization of Enzymes Encoded

[00956] *Objective*: To identify additional triterpenoid aglycone biosynthesis genes in the triterpenoid saponin biosynthetic pathway.

[00957] *Methods*: See Materials and Methods above.

[00958] *Results*: SOAP1, SOAP2 and SOAP2-like nucleotides (nucleotide sequences SEQ ID NOs: 45-47 were expressed were subsequently used as baits in co-expression analysis ($r > 0.9$; Pearson correlation coefficient - PCC) and an additional five genes encoding p450 cytochromes, eight genes encoding glycosyltransferases, and five genes encoding acyltransferases were identified for functional characterization using VIGS (**FIGURES 20A and 20B**; **Table 14**). [SOAP1, Soap2, and SOAP2-like nucleotide encode polypeptide sequences: SEQ ID NO: 48, SEQ ID NO: 49, and SEQ ID NO: 50, respectively) These experiments revealed that CYP72A655 (SOAP3; Sp_085340_meek; SEQ ID NO: 51 [nucleotide], SEQ ID NO: 52 [polypeptide]) and CYP72A654 (SOAP4; Sp_040350_wdny; SEQ ID NO: 53 [nucleotide], SEQ ID NO: 54 [polypeptide]) participate in the biosynthesis of the Yossoside aglycone (**FIGURE 25**). In **Table 14** below, terms used include: coexpressed gene - gene with coexpression correlation coefficient greater than 0.9; annotation - gene functional annotation based on blast analysis against *Arabidopsis* proteome (TAIR10); and aa seq - amino acid sequence of coexpressed gene

[00959] **Table 14: Coexpression analysis of spinach genes with three baits SOAP1, SOAP2 and SOAP2-like.**

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
3	sp_000460_nyyz	0.984337	0.95971	0.970205	Pentatricopeptide repeat (PPR) superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT5G01110.1); Has 32626 Blast hits to 11011 proteins in 270 species: Archae - 3; Bacteria - 29; Metazoa - 476; Fungi - 329; Plants - 30874; Viruses - 0; Other Eukaryotes - 915 (source: NCBI BLink).	MVSFLRHSIPAVTNLFLSIGKPTKRLLSHRLVADD PKHTVASSVCDSLRKGLNWDTLSRKYGSVNLDD LIIERVLLEFKEPIEAKRALSFFHWAAHNNKLGHG LKHYCIIIQILARARLVNDARSLLESILMRNGENEV NCSLRFPVVDMLLNTYKVTNSCSFVFDLMVQSY SKLRMYSAAFDVCCYLDERGFSLSLVTFNTLLHVI QKSPDDSNLIWEVYGHMIQKRINPNDATVKIMINA LCKAGELQKYMNIIGRVHGKGCSPMVTVNTSLVF RMLGEGRIEKGLVLLKRMLQNKLLLDDVSYSLVV YARLRSGELNLAREVYEQMLSSGFLPNSFVYTLF IEGYCEKGMVEEALNLMDEMVNMGLKPYDDTYN SLIVACARIGASVDSFRLCEEMTEKGLLPSTSAFN EMVGKLCESGNMKQSNDMLTVLMEKGYLPDET TFLHLIDGYGREKNIQEALKLYYEMEFRSISPGLPI LTSLIRTLCKCRKLEEADKYLAIMKRRSLSLTVCIY KPLISAHLENGNKERAHHLYNEMIQRGVKLNQPM PPDFFSEDLDS
4	sp_000760_grig	0.940658	0.963699	0.991838	HTR12; FUNCTIONS IN: DNA binding; INVOLVED IN: double fertilization forming a zygote and endosperm; LOCATED IN: chromosome, centromeric region, nucleus; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Histone H3 (InterPro:IPR000164), Histone-fold (InterPro:IPR009072), Histone core (InterPro:IPR007125); BEST Arabidopsis thaliana protein match is: male-gamete-specific histone H3 (TAIR:AT1G19890.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MRTKHPAAKKSSSKSPRSKAPSTPSKSNQSKST PIRSPKSKSGRKSVAAPRIKKPYRHRPGTVALREI RKYQKSVDLLIPSAPFVRTVKEISHQFSPMVTRW KVEAMKALQQAAEDFIVCLFEDAMLCAFHAKRVT LMKKDIELARRIGGRERRW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
5	sp_000910_wmxj	0.98544	0.965034	0.981735	gamma-glutamyl hydrolase 2 (GGH2); FUNCTIONS IN: omega peptidase activity; INVOLVED IN: tetrahydrofolylpolyglutamate metabolic process; LOCATED IN: vacuole; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase C26, gamma-glutamyl hydrolase (InterPro:IPR015527), Peptidase C26 (InterPro:IPR011697); BEST Arabidopsis thaliana protein match is: gamma-glutamyl hydrolase 1 (TAIR:AT1G78660.2); Has 406 Blast hits to 402 proteins in 115 species: Archae - 3; Bacteria - 51; Metazoa - 191; Fungi - 0; Plants - 77; Viruses - 0; Other Eukaryotes - 84 (source: NCBI BLink).	MFPSNTPSSTSHLHSNLWFFLWIPLITSSLSQQL TSAEPNNNIIMLPTSSDYLPAPKSCPVPDPDLNFR PVIGIVSHPGDGASGRLKNGTNVSYIAASYVKFV ESAGARVIPLVYTDPIDVIFKKLELVNGVLFTGGW AKNGTYFEVVQQIFKKVLEKNDAGKFFPCYAICL GFELLTMIVSKDNNILENFSASYQASTLQFVDNNV SGTVFQRFPPELLRKLSTDCLVMQNHKFAISPKT LQANHLLSSFFDVITTSTDSDNKVYVSTVQAHQY PITGFQWHPEKNAFEWGYPMIPHSEDAIQVTQH VANFFISEARRSPNRPPVPDVLNNLIYNYSPTYC GKKGNGYDEVYIFG
6	sp_002040_oxuo	0.940411	0.973901	0.978911	purple acid phosphatase 29 (PAP29); FUNCTIONS IN: protein serine/threonine phosphatase activity, acid phosphatase activity; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Metallophosphoesterase (InterPro:IPR004843), Phosphoesterase At2g46880 (InterPro:IPR011230); BEST Arabidopsis thaliana protein match is: purple acid phosphatase 14 (TAIR:AT2G46880.1); Has 636 Blast hits to 631 proteins in 208 species: Archae - 0; Bacteria - 277; Metazoa - 0; Fungi - 208; Plants - 117; Viruses - 0; Other Eukaryotes - 34 (source: NCBI BLink).	MHYANGKDTPCSDVLPSQLQTCSDLNTTDYLRR FIQTENPDLVVFTGDNVHSTDAYNYNRSMDEVFA PAIESNIPWTAILGNHDVNVLQTGIPRQAVMEYIV GMKNTLSLLNPSNLEEDELIWGWGNYNLEVAGI KGSPFENKSVLNLYLLDTGDYTYSFTPTFYLGYE WIQPSQQSWFLRTSKQLQETYMSAPQPQPQQG PAPGLVFFHIPIPEFNNISKSDMVGVKQENVICPG VNSGFFDTMLAAGDIKAAFVGHGHTNDYCGKLC DIQLCYDAGFGYHGYGIAGWPRRARVINVKLDKT QDGNWGPVNSIVTWKRLHDDRLTIIDPEVLWTS AGRDQSKSVFKPIMSE
7	sp_005360_psus	0.960852	0.94178	0.96636	DNA/RNA-binding protein Kin17, conserved region; FUNCTIONS IN: zinc ion binding; INVOLVED IN: biological_process unknown; LOCATED IN: intracellular; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), DNA/RNA-binding protein Kin17, conserved region (InterPro:IPR019447), KOW (InterPro:IPR005824); BEST Arabidopsis thaliana protein match is: DNA/RNA-binding protein Kin17, conserved region (TAIR:AT5G51795.1); Has 1885 Blast hits to 1561 proteins in 251 species: Archae - 4; Bacteria - 25; Metazoa - 715; Fungi - 242; Plants - 160; Viruses - 13; Other Eukaryotes - 726 (source: NCBI BLink).	MEGKSGLKALSNRMKSKGLQKLMHYCEMCQKQ CRDANGFKCHCMSDGHLRQMQIFGENPKKFIEL FSQEFEDAFLELMRRSHRFSRVAATVVYNEYISD RTHVHMNSTKWLTLTEFVMHLGRTGKCRVDHTE KGLFLTYSARDSENLFYERRKNKRLKSVLAEEVK QENEIVKQIERAAASVAVPVETEVRTPLELQINNI MLGSEPEMGKNKKISFAFGFGSSSTSVTSLKPRE DSSMSFGLEDVPNAELGLSKKKTIGDVDGSGGC VS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
8	sp_008840_kejm	0.98056	0.951213	0.97876	receptor-like protein kinase 1 (RLK1); FUNCTIONS IN: in 6 functions; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480), Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase-like domain (InterPro:IPR011009), Protein kinase, catalytic domain (InterPro:IPR000719), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: lectin protein kinase family protein (TAIR:AT1G34300.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MASPPSCILFSFFTLFLLQSIAAQNGGQITVGQSL TAIGNGSLLLSPSGDFAFGFHQLPNKNNLFLLAIW YAKIPDTIIWHANDGNPVPQENICMTDQNPRPEL GSGICGFNSICSLGDASDDRRPTCTCPQGYSFR DPNDKYSSCKPDFLDGCEGYAERSLKGEYNLKL VPTTDWPFNDYERLDPYNEEDCKTSCLKECFCA AVIFKEPNTKDHGCWKKRAPLSNGRKDTTVLGT TWIKVGNVNISNDPLNPFVSDPPSKAKNKVKSLT KVLLGGQRKTVTEYKNIDCFSYQELIEATNGFKE ELGRGAFGVVYKGMINGGGLSTSVAVKKLDRISG DTDKEFKTEVNVIGQTHHKNLVQLVGFCKEDDQ RLLVYEYMSNGSLAAYLFDGLRPSWKDRVKVTQ GTARGLLYLHEECSTQILHCDINPQNILLDEHQNP RISDFGLAKILVLNQTHTNTGIRGTKGYVAPEWF RNKPVTAKVDVYSFGVLLLEIICCRESVCMELIAE KEPF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
9	sp_008960_semr	0.9884422	0.8833478	0.9358831	receptor-like protein kinase 1 (RLK1); FUNCTIONS IN: in 6 functions; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480), Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like comain (InterPro:IPR017442), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase-like domain (InterPro:IPR011009), Protein kinase, catalytic domain (InterPro:IPR000719), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: lectin protein kinase family protein (TAIR:AT1G34300.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAKPFTCIFSFFATLIFLPLLPIYAQSSGKIAVGQS LSAATNASPWLSPSGDFAFGFHQLPTNSNLFLLA IWYAKIPDTIWYANDGNPVREGSKVTLTADKGL VLSDPQASSLLWSTSDDFSNGNGVSYGFMSDT GNLVLKNSSSNDPVWQSFNHPTDTLLPTQIMEIN GVVNSRLSENNFRKGRFQLRLIPNGNLVLNTRDR ASGYNYGAYYISGTNERPNTGHRLIYNESGYMYV LRTDGSTYDLLPQNKTNPTKSYYQRATLNFDGVL MWYSIPKTSISNGWSMDQALPDNVCTSIGASGP QLGSGICGFNSICVLGDDKRPTCKCPPGYSLLDS NDTYGSCKPDFKADGCVEYAQRSMKNEYRLERL PSTNWPFSDYEKLGPYNEDDCKMSCLTDCFCAA VAFQDSNSPICWKKIKAPLSYGRQDTKVRETTWI KVGNVNFSNDPLNTFVPLFPPKAKNKVKSLDKFL LGGSVFVNLVLLSAISLGIFLIYHKKPLKEFEEGQR KKLREYNSNVHCFSYQELTDATKGFTEELGRGA FGVVYRGMISTVGPPICVAVKKLDRISHDADKEF RTEVNGIGQTHHRNLVRLVGFCKEGDQRLLVYE YMSNGTLANYLFGELRPSWVARIQITLGIARGLLY LHEECSTQIIHCDIKPQNILLDDNHNARIADFGLAK LLVLNQTHTNTAIRGTKGYVAPEWFRNKPVTVKV DVYSFGVLLLEIICCRRSVCMEFIEEEGAILTDWA FDCYQGNTLDSLVNNDMEALNDWQRLKRFFMVS LWCIQEDPSLRPTMKMVMQMLEGLAEVPKPPCP TMLSITTTNTKPLG	
10	sp_009290_wckf	0.9885833	0.8611586	0.8372293	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G25265.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MMMNGSKIMGRGGMVLFVALFIGFCFAIHYSSTL SLTNTQLKVLGKYGPARQVELVVETPVDPKQDA VSGPTRMKFHVAVTSTDSIYSKWQCRIMYYWYK RVKDLPGSDMGGFTRVLHSGQPDNLMEEIPSFV VDPLPEGVDKGYIVLNRPWAFVQWLEKATIEEEY ILMGEPDHLFVNPLPNLAQGENPAAYQPFYIRPD LNEKIVRKFYPNGSISNVDPIGSSPVIITKSLLKEIA PTWMNISLRIKNDPEADKTFDWVQEMVAYAIASA LHGVKHKLHEHFIVQPPMDGNVEISYIIHYTYGCD YNNKGELLYGGQKTDWRFDKRSYQDAPPPKNLTL PPLGAPETVVRLIQMVNEATANIPEW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
11	sp_010800_skmm	0.947733	0.961332	0.986633	SUPPRESSOR OF ACTIN 9 (SAC9); FUNCTIONS IN: inositol or phosphatidylinositol phosphatase activity; INVOLVED IN: response to high light intensity, phosphatidylinositol metabolic process, response to cold, phosphoinositide-mediated signaling, response to osmotic stress; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: WW/Rsp5/WWP (InterPro:IPR001202), Synaptojanin, N-terminal (InterPro:IPR002013); BEST Arabidopsis thaliana protein match is: Phosphoinositide phosphatase family protein (TAIR:AT1G17340.1); Has 3998 Blast hits to 2752 proteins in 237 species: Archae - 0; Bacteria - 0; Metazoa - 2445; Fungi - 904; Plants - 298; Viruses - 0; Other Eukaryotes - 351 (source: NCBI BLink).	MEAMKLEIERLRLNLSTAERYRALLAVSTDPASIN PNLLLDDLQMIRFCRITSLAMVGQAALEDKITKSI GLEKNDDDVIDFWNVSGFRQSCVGGVCEVHAE NGGGGANIYREIISRFLKVFVVFSLWEKRLSNML CWKGSISACKL
12	sp_010890_uhjp	0.963735	0.965334	0.983931	uncharacterized protein	NKVQVPEVQTEVYHVAWPDQEKEYVPETEFVLS TEKVNDQPFPQGVSSLCSLAIENDKGVHIVAVGE VAIIKFSAEYRKDIPKIKKNPFIKWQKIECPRQPPG SKESGYYVGRYMIETIALRQMFIREKTMLFLLLRC CYAAVPAADSDVPAAGAAVAGEFVRCCC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
13	sp_011580_zgon	0.962074	0.938467	0.982534	Leucine-rich repeat transmembrane protein kinase; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Serine/threonine-protein kinase domain (InterPro:IPR002290), Leucine-rich repeat (InterPro:IPR001611), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Malectin/receptor-like protein kinase (InterPro:IPR021720), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat transmembrane protein kinase (TAIR:AT1G53430.1); Has 190921 Blast hits to 136069 proteins in 4848 species: Archae - 114; Bacteria - 18757; Metazoa - 52699; Fungi - 11122; Plants - 85531; Viruses - 458; Other Eukaryotes - 22240 (source: NCBI BLink).	MAPFFFTSLLFFQLTSLFIIFFTNFGSDAQLIPQIEV KTLEIISQKLENTYWKVTQNSCSDGGRSFNISIGV NILSNVTCDCSFNSGKICHVTHIQLKGLNLTGSLP EEFGDLTQLQEIDLSRNYVNGTIPKSLSKIPLVTLS LLGNRNIGPIPTEIGQISTLEELILEDNQLGGPLPK SLGNLKSLRRMVLAANFFNGTIPETFVNLKNLTEL RLSGNVLSGKIPDFLGNLTNLNSIDIQGTSLEGPIP STISSLTNLTILRISDINTRSFRITFPNLTNMANLQY LVLRNCSITGSIPDVIGTLANLKLMDLSYNQLSGQI PQSMENLRRLENLFLTNNSFSGTLPSWILNSRHN FDVSYNNFTGSSPASCQQSNVNLVASYSSAKSN SVNWCLRRDLPCSNKPQYYSLFINCGGTKLSFQ GNNYDEDSSPGSPSTFFADGEKWAYSSTGAFLY DDNGNFLAMETSNHNVTGIYQTARLAPLSLKYYG LCLLQGSYKVKLHFAEIMFSDGQNFGNIGRRFFD VAIQGDVKLKDFNIESEAGGVGKAVVKEFDVNVT GSTLEIYLYWAGRGTTGSPDRGVYGPLISGIAVT PNFKVNQGLSAGAIVGIVLASCAILVLVLVVLWKR GYFGGKSEENEELRRLGTGYFTLRQIKTATDDFD HRNKIGEGGFGPVFKGVLPDGKVIAVKQLSSKSK QGNREFVNEIGMISALQHPNLVKLYGCCIEGKELL LVYEYMENNSLARALFGKENQKLHLDWQTRKRI CLGITRGLAYLHEESRLKIVHRDIKATNVLLDKDL NAKISDFGLAKLDEDENTHISTRIAGTMPKEEFVY LLDWAYVLQEQGNLLELVDPTLESSYPKDEALRL LNIALLCTNPSPSLRPAMSSVTSMIEGQTPIQAPI VKRSGANDDLRFKAFERLSQDSQTHVSGSSMSS QVQRSISIDGLWSDTSVSMQSKDGITEMSVSTSR LLPDDVKLE
14	sp_013360_axec	0.975674	0.924591	0.92219	RAD-like 6 (RL6); CONTAINS InterPro DOMAIN/s: SANT, DNA-binding (InterPro:IPR001005), Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), Homeodomain-like (InterPro:IPR009057); BEST Arabidopsis thaliana protein match is: RAD-like 5 (TAIR:AT1G19510.1); Has 590 Blast hits to 589 proteins in 66 species: Archae - 0; Bacteria - 0; Metazoa - 121; Fungi - 0; Plants - 466; Viruses - 0; Other Eukaryotes - 3 (source: NCBI BLink).	MSSNGWTPKQNKAFERALDKYDKDTPDRWNNI AKEVGGKTVEEVKKRYQDLLEDIQHIESGRVPFP YNWKM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
15	sp_013520_aqcr	0.936704	0.931229	0.93759	ELMO/CED-12 family protein; INVOLVED IN: phagocytosis; LOCATED IN: cytoskeleton; CONTAINS InterPro DOMAIN/s: Engulfment/cell motility, ELMO (InterPro:IPR006816); BEST Arabidopsis thaliana protein match is: ELMO/CED-12 family protein (TAIR:AT2G44770.1); Has 873 Blast hits to 873 proteins in 155 species: Archae - 0; Bacteria - 0; Metazoa - 490; Fungi - 72; Plants - 180; Viruses - 0; Other Eukaryotes - 131 (source: NCBI BLink).	MKIYFDPSRSDHQEALKVLWYTTYPDQPLHGLIS EQWKEMGWQGRDPSTDFRGAGFISLENLLFFAK TFSTSFQHLVKKLGGKRATWEYPFAVAGVNITFM LIQMLDLDAPKGRTFVRAVFLQMLSDNEWAFDLL YCVAFVVMDKQWLEKNATYMEFNDILKSTRAQL EKELLMDEVLGIEDMPSFSLLC
16	sp_016480_idkw	0.925769	0.973315	0.945384	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT3G57430.1); Has 47794 Blast hits to 13382 proteins in 245 species: Archae - 0; Bacteria - 5; Metazoa - 44; Fungi - 66; Plants - 47218; Viruses - 0; Other Eukaryotes - 461 (source: NCBI BLink).	MFNHVRHLPLISPFSNPKPLSFSLLSTSAIKALPS SPHHFLSPNSIDFLPLLSSLQSLHQTKQSHALALV FGFLPSSVSLCASLILNYAKFGNPVSCKSLFEKSV FYCKTPFLWNTLLRGYSNSGVYDVFYVYNHMIRT HVRPDDHTFPCVIKACADHCENKKGREIHGFSIK VGFDNLVYVGNTLLSFYSSCCDLVGAQNVFDEM PDRDVVSWNTMIGVFSANGWSNEAVEVFLQMKI WSEFLPNAVSVVSVLPACAELKDEVVARVIHGYV VKVGLINQVRVGNALVDGYGKCGELCSSKQVFN EIIEKNGVSWNTVISSFAHAKCYSDALYMFMLMID SKEKPNGITLSSLLPILVELECFDAGKEVHGLSIKT GLDSDVFISNSLIDMYAKARNFTKASIIFQNTSVKN VVTWNAMVANLAQNSQELAAIGLVRKMQSFGEE PNAVTFTNVLPACARIGALLSGKEIHGRCIRKGIS PDIFLSNSLIDMYAKCGLVDLAQNVFNVSNRDQIS YNTLIVGYSQTTECLKSLTLFKEMGLIGMEMDTIS FVGAISACANSAALKPGKEIHAFITRRLFHTHSFIS NSLLDLYMKCGKIGLAQNIFNRAHKKDVASWNT MILGYGMVGDFETAISMFEAMREDGVKHDSVSFI AVLSACSHGGLVEKGRKYFEEMKSLSIVVKNMH YTCMIDLLGRAGLMEEAVELIKELPFEPDANIWG ALLGASRIHGNIEVAKLAAENLFVLKPEHSGYYVL LSNMLAEAGRWDEATRVRELMKSRGVKKDPAY SWVQNRDQVRAFLVGES

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
17	sp_017870_ceny	0.985764	0.931744	0.958512	Phosphoglycerate mutase family protein; FUNCTIONS IN: catalytic activity; INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Histidine phosphatase superfamily, clade-1 (InterPro:IPR013078), Phosphoglycerate/bisphosphoglycerate mutase, active site (InterPro:IPR001345); BEST Arabidopsis thaliana protein match is: Phosphoglycerate mutase family protein (TAIR:AT5G04120.1); Has 16927 Blast hits to 16736 proteins in 2503 species: Archae - 87; Bacteria - 12791; Metazoa - 870; Fungi - 486; Plants - 272; Viruses - 0; Other Eukaryotes - 2421 (source: NCBI BLink).	MAASDSRVSMECVKNDTIEGAAEPVDGTCTEIIV VRHGETEWNALRKMQGQMDIDLNEVGRQQAVT VAHRLSRETNISCIYSSDLKRAFETAETIASRCGG LEVVTDKDLRERHIGDLQGLTFREAPMVNPIAFE ALKSPENDQEIPGGGESRDQVYKRCTSSLQRIAE KHRGARVVAVTHGAVIEMLYKRAIPGGRAEGIWN TSISIFQLFEGDKWSIKLWNDKSHLRETEYLESAS GGGDASC
18	sp_020820_yeau	0.977992	0.949386	0.98603	UDP-glucose:flavonoid 3-o-glucosyltransferase (UF3GT); FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: N-terminal protein myristoylation, response to sucrose stimulus; LOCATED IN: cellular_component unknown; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, C globular stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT5G54010.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGKTVAASEQLHIVMIPWFAYGHILPYFELSNKLA EKGHKITLVVPNKVKLDLEPKIRHPSLISLHAFTVP HIEPLPPGTETCSDVPIELQHHLAVAMDRARPEV ESIISAIDDPKPDLLFYDNAYWVPEIATKLGMKSV FYQIACALSITRIKQTPSASASASASAKLFTLPKW VLTPKVLGDARANYGEGITYYQRVKKALSSCDAI ALRTCREIEGESSDILAAQYNKPVFLTGPVLPEVE FLPPLDNSWAEWLAKFGPKSVVLCCFGSQYVPD KAQLQEMALALEDTGLPFLMSVKPPTECATIEEA LPEGFSERVKERGVVHGGWVQQLQILAHPSVGC FICHCGYGSMWEGLLSDNQLVLLPQLPDQLMMA QMLAEKLKVGVMVDREEDDGWVSRKNLCQAVK SVMDPHSEFAALLKNNHANFRDKLLTNGFMANY LEVFDQDLKRFLTAN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MWYKIEGIFTPQNNFNVAFHALERVGRVIRFDET
						SRKEGPKEFLRAKVIIPTRRPLIPGFYYEYSQDNF
						KWVHLRYEGVFNFCQNCGKIGHRKPNCQVQEE
						LVKMEVRHRLKGICGLNFEFMEVTDAPPLYSKKV
						MGLDDIEKFRTTFRLNLRRWGQPTATDSDNNNDS
						SDSYSPPSPPDNPSQRSTGSNSATSATNKTNHT
						GKKRKGDNQAQGGGSNSKKRSRRANGEASDLR
						SIRERRGLANLQKLKLPSQNILMGWCMADSVPE
						NGSISKAVAGLVTPSQPCKNRVSHYHQSSSNSL
						FNASGHTINTSCNLESASLLTIPQLQSPTWVESS
						PRLLVELPVFSIKWSVPYPSFIMTAPTVFPPSFW
						GFFLENFVLYHSPVFLSSPVYSLEVELEFKLPVLM
						ISTTCTSIDSLLNLQHCYLYPNTFSLNLHKPYPIPQ
						TILVPHPDLLPIFPLLPLLAMDTPPPMSTSLVIASP
						KAVVLSQQDRIFTPKNTLLDPFKPKSTSSAYSWS
						YNTNIQFGSIPKHLVSAQLSPKFANNSAGPSSTPI
						ETLIHACSQNLQLALSVPLSPVNGDTSPNVSTPP
						SSPASVNSFITAPSSLESAEVESMIKLDYLALDSD
					pleiotropic drug resistance 11 (PDR11); FUNCTIONS IN: ATPase activity, coupled to transmembrane movement of substances; INVOLVED IN: drug transmembrane transport; LOCATED IN: membrane; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: ATPase, AAA+ type, core (InterPro:IPR003593), ABC transporter-like (InterPro:IPR003439), Plant PDR ABC transporter associated (InterPro:IPR013581), ABC-2 type transporter (InterPro:IPR013525); BEST Arabidopsis thaliana protein match is: pleiotropic drug resistance 6 (TAIR:AT2G36380.1); Has 357151 Blast hits to 273757 proteins in 3981 species: Archae - 7090; Bacteria - 287486; Metazoa - 9208; Fungi - 6204; Plants - 5777; Viruses - 4; Other Eukaryotes - 41382 (source: NCBI BLink).	STTMNSWDVEPEYYSHEVKKMIYKKRAGMRQH
						EDGPRKKKWCPSLDDEIRVAIVEGKIYPEEENPII
						LKRIIEGQGTSGNHGANSSNSSSPWKKKKKAQLR
						YEGRGKFRFGLSKDDDLWEKLWVVAEEGGEDINT
						NKTSINILEPGSKLFKVQRKLQQVLNKTRKWCLD
						YKKEMGITWNKFEEELAIIQGNIHNIESSIGEVEER
						KKCREKAEEQWQYWRQRSKSKWDAFGDQSTA
						FFYKSVKGRSCRNDIRAIKDDDGVWKTKKEEIQN
						EFYKYYNNLFAGSPYVAGHEDNMKFWLEDIPSIIT
						AHINALMQPFSEEDIKVAAFASKPLKSPGPDGVP
						PVFIQKYWATAILGFIRLAPSKKRTLDILHQVSGIIK
						PSRMTLLLGPPSSGKTTLLLALAGKLDNNLKVSG
						KITYCGHELKEFVPKRTSAYIRQNDIHLGEITVRET
19	sp_020990_kogi	0.873978	0.952413	0.987857	(source: NCBI BLink).	LDFSGRCLGVGSRYLLLKELLKQEKQAGINPDVE
						DTMARKKKTVNLADLRLMGHQMDFSFSGGFMFT
						MSAGEISRTFPGEAKSSHHEATYFDEFWMEETE
						DVGPSNHLVDSLEESEQDESDGGVEEEAAERPR
					GAMMA-IRRADIATION AND MITOMYCIN C INDUCED 1 (GMI1); BEST Arabidopsis thaliana protein match is: defective in meristem silencing 3 (TAIR:AT3G49250.1); Has 114 Blast hits to 102 proteins in 24 species: Archae - 0; Bacteria - 7; Metazoa - 8; Fungi - 4; Plants - 91; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	VAAVVRGENDQEMELKCSRWQLEPHFAGGSDE
						LPALTKQEWDQIVVHFQVESLAVQGKNLNIPKS
						WLPSCSQFKDHVFLASAGLDHGLDQAFAMSKFP
						PVEKGQINTKEWLTQMVEQKLKGIVGGPPAKFE
						FRCGCISIPRRICKQLSGGIFPLVQVDSDGFIDLS
						SLLKVTEGYGKPVSLSVKYNKTVIANKKFLVDKR
20	sp_021210_xeed	0.945435	0.924177	0.9781		ELRMVTEVQS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
21	sp_023090_ietn	0.951575	0.967691	0.99026	AGAMOUS-like 20 (AGL20); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: flower development, response to cold, positive regulation of flower development, maintenance of inflorescence meristem identity; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Transcription factor, MADS-box (InterPro:IPR002100), Transcription factor, K-box (InterPro:IPR002487); BEST Arabidopsis thaliana protein match is: AGAMOUS-like 42 (TAIR:AT5G62165.3); Has 7364 Blast hits to 7361 proteins in 926 species: Archae - 0; Bacteria - 16; Metazoa - 673; Fungi - 316; Plants - 6262; Viruses - 0; Other Eukaryotes - 97 (source: NCBI BLink).	MVRGKTQMKRIENDSSRQVTFSKRRNGLLKKAF ELSVLCDAEVALIIFSPTGKLYEFASSRCCSTSRTI DRYQKHMKELRGVKSKVSDRNMLQFKEDANELE KKLKSLELSKRRLLGDSLESCSIDELQDLEKQLDK SLSTIRAKKNALFKVQIDKLYEEEKLLLKENARLKE KLQNSLPVPLSVRQPTDDDQRNNVQDDVETELFI GLPESRNSRCLP
22	sp_026540_srmp	0.954322	0.944109	0.970126	uncharacterized protein	MRWSSARAHVVDMTSVVDGSGVVAGQVSSAG QVSAIG
23	sp_028080_hieu	0.982014	0.994455	0.995851	NB-ARC domain-containing disease resistance protein; FUNCTIONS IN: protein binding; INVOLVED IN: apoptosis, defense response; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing protein (InterPro:IPR015766), NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: LRR and NB-ARC domains-containing disease resistance protein (TAIR:AT3G14460.1); Has 25652 Blast hits to 19792 proteins in 681 species: Archae - 38; Bacteria - 1467; Metazoa - 4997; Fungi - 304; Plants - 18213; Viruses - 0; Other Eukaryotes - 633 (source: NCBI BLink).	MAESAVLPLLAEAAKPMIGVLVSRVTAEIQLCRDF KRDLSELQRKIQIISSTVLNMTSATNTRSSSSSSN TMELWVRDVVNIVHNVEDLLDVLAYEKQRQVVM KRTFLWPLKQVYFRRKISRWIKRINTITGEALIEA QTIIQLKDLVHSTSPLPTATNFITNEGAAAAAENR LQQLRRYVDISLIVGRDSDVDSVVDKLCNPSKDC YNTSNNCLSVVAVVGIGGIGKTVLAKKGSPGGLL KHVKYLRVLDLCDMGLEEVPESVAGLKCLKSLSL AANPIKVLPEFVTELYNLQSLNLLKCDLLKEVPGG LSNLVNLRHLMIKILNLPAGMIGQLRCLQTLPPLW VNEGGFQISELGDLPHVSGSLAIHGLEHINSKQD AESASLSEKSRVKKLRLLWDSDDDWIKHKEVLD GLKPKSNLRYLIICGYGGEIFPSWLMRMDVIKIIEL VGCRRCQKLPTLGHLPFLEKLLIEDMESLECIGEE LHCNTAILTEVKSSMVYFPSLKQLRLIGLHKLTEW VEPSSGGWTLFPLLEELKMAYCHELTTTPNTFPS LKQLTVTNITSGRPLSNIITNKSSVKCLKVLQLDWI TDITDLPEYLAESCKSLQTLKIGNCVNLSFLPENL SMLISLQVLEIKRCDKLVSIPNLSNLGSLRELMIFH NLALTSVPKGLQSCNSLERLWFACCPKVETLPDL TGLTRLCNLSIRSMTKMITQPPDWLYTLPCLRLL GIGSYWELDEVPDMSFLVKIVSLESLHLFGWEKL QSLPEQLQQFTTLKELGICDCDNMEYLPDWVGN LSSLRYLSPTRCRNLKRLPTMDAMKRLTKLSFLKI EDCPELEEAVQFNGSEYHKISHIEIKIVYPTY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
24	sp_032370_ufri	0.989628	0.965555	0.953182	unknown protein; Has 109 Blast hits to 109 proteins in 44 species: Archae - 0; Bacteria - 2; Metazoa - 61; Fungi - 8; Plants - 36; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MVKGLQQQNLPADVTQLVDQLERHCLASDGSFV SKSAYYDLQLAREEMCRERLRYLEAMAIYCEAIA MVEEYQQALSVANRDVQGLYPQLGSKNSPQVY ETLEHRLVVAEAAQKLRLPLISKDGEIHEEDIEKW SIMSRSSLDSTSTNLTMSSSTNSTNYGTSTPSTV ANVTPSNSTDMSDPGIGGVPNRYLGITSTYLWQ TQLHQGTAEYQMSLFREIEARLRVKCDQLADAFV MDDVDGSSGYQNVSSRLPERVKIIIEEIERDEEAL KEDLYSADRKFAEYYNVLEQILGVLIKLVKDLKLQ NQHKYDELQKTWQIKRHEAMGAKLRVLDHILLVA TYTQETIPALHKIRNYLVEATEEASTAYNKAVTRL REYQGVDPHFDTIARQYHDIVKKLENMQWTIHQV EMDLKYLPDHSST
25	sp_033550_oqaz	0.937021	0.955157	0.950556	RESISTANT TO P. SYRINGAE 5 (RPS5); CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing protein (InterPro:IPR015766), NB-ARC (InterPro:IPR002182), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: LRR and NB-ARC domains-containing disease resistance protein (TAIR:AT1G12280.1).	MLANEVITLGAAIFGLLTSIAGSTACFYNSYKTSRI NLSILKVKVQNLEILIQELQQKLNGIDVLKHDPRE WISRAREFLETVNISDLESRLNENCVGGQLPNFC SRFGALKSIEESNHGLDVLLDSGKNVSREIKVKG SIQPEVNIVGNARKEIKKRIWEQVSQAETAGIVCI HGMTGVGKTAFAAAIHNKALKELTAFENVIWVNV ENGSGLKHIQEELASKLQVHLPGTTVNKETLRSA LMQRGRFLLVLDSMWQAFSLHDIGIPEPVAGSKL IVTSKTYSVCRKITKNIKSREIYEIKPLLEAEAWDL FQGEVGPNIVNLESETVSIAQRAVKNLDGLPLAIK LFAETLSEVHDDYGSGIDAAWKDELFSLSRSTSL LEGYLIKPKELIDYWMWEGLLGPISSLGESMRLG RQTLNELKFAHLLESVKSESGEESLKMINLVRHV AMDAISKSGNHFFIKCGQSLGEFPSAYDWPANT ERASLVQNQLRALQNSPDCNKLSTLLLQDNPLNL VHHENFFSKIRNLRVLDLSRTSISSLPKSFSGLTN LHALLLRSCRNLKLLPSLSNLHRLIVIDISETPLEQ VPDGLHNLINLRRLDLSKTKIEVFPATVLGKLTEM EELLLITAENGGYVWGSNLIFPQWPGACIEELLDL KKMVVLQLAFLNPEVFNRFVDQVDDQQRAIQLR SFKFYVGGIYSGGDIGDNSVVVMGDCSIRVPRGT SELYLMKHSQELLSLKFDGCMRDLTIVDVSRFDA LTYLFTLEMLGSLKSLKKICVKHCKKMKSVIQPAA EANPTSITIPKLTNFKLYDLQNLESIYDGKVLSCPS LHGFEVFNCKKLKLPKVLVGRQNCIIEIVGEKEWI DKVMEDGPSPHLKYTEALVPAELSTPSFRLRRSE SFVSFRDTHEDSGRSLGATIPRWHRLFHPFRGS AVFSPAADTLHIASTTSNDAISEVSVLPHAEAGTS QRGFNRTRSSRSD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
26	sp_037920_hhky	0.955715	0.953679	0.93337	unknown protein; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF2439 (InterPro:IPR018838); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G43722.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MDDVKKWTVTYTKHVKQKRKVYQDGSLHLLSNN KVMLYDESEQLLDSKFLTKDEAVKPGESLTFASF LVDIGDFDNNCNPTSNANPKPRDKTQMPQREKS RHTFSSMKPNTAGRKTSLVSVSPSQKIIREFKKN EELKYGAHSPKSPGTSSFKEWEALYTTQLTQKA KKYHDGFIRMIYCGSQGRQIMLYDSSKNQIAKRF MKKDEVIDTGGSLTFDSHLVDIGDPMGTNGDDA DVKIQERNNAKRSKKKSTLLDHVGSGNSGVIGKRR DDACSTKRVNSDSGSFGFDKINQKKNFASSSSL RDANQILSILRKPITQDTISAVEEAPMQQCPPTSS EPVQIHADSQAFDSSLQVDQCPVIKAVEKHNNEK AVDGGQSSTFSEGNCSYLERLCTQGIEILRSNEF ESPPKSQPIVGSSVDHGRPIVIQSSKRVESIDNRS VVDSETTIHLQPSTTKLPETDPENKEKTASTKDM SKRFGSLMRMKTKKSVKNEKLNIKMECPSFDLGI
27	sp_040350_wcny	0.9573027	0.9643238	0.9881317	cytochrome P450, family 72, subfamily A, polypeptide 15 (CYP72A15); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: response to cyclopentenone; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 72, subfamily A, polypeptide 13 (TAIR:AT3G14660.1); Has 31875 Blast hits to 31741 proteins in 1631 species: Archae - 65; Bacteria - 4943; Metazoa - 11298; Fungi - 6149; Plants - 8001; Viruses - 3; Other Eukaryotes - 1416 (source: NCBI BLink).	MISKSASGVAIASLGCIILYWMWKLLKGLWLTPK KLEKCLKQQGLVGNSYKFLIGDMKESSKLRNEAL QKPIPFTHDYYNRIQPFIHQILNNSGAGKNIYTWL GPVPTILITQPELIKDAFNRMNNFQKPRLNPYTQM LSTGLPNYEGQKWAKHRKLLNPAFQLDKLKLMIH TFETCVTDTLNKWEKLVCKTGSSEVDIWPYLTTL TGDGIARAAFGSSFEDGRRIFELLTLQKDIVISLLK YSYIPGFKYMPIKGNRKMKEADNEIKPLLTNIINRR RKAMEAGEAPKDDLLGMLLESNANEARQVNENE SGSSKRKSDLTMSFPEMIDACKQFFLAGQETTSV ALTWTMLLLAKHQDWQTRARQEVLATFGMNTP DFDGIHNRLKIVTMILYEVLRLYPPVPATSRRVHD RETKLGDLVIPQGVGVSFSILHAHLNPEIWGDDA KEFKPDRFAEGIAKATKGNNSYFPFGWGPRICIG QNFALVEAKMALCMILQRFSFDLSPSYIHAPTSLI SLQPQHGAHILHRF
28	sp_044310_ajio	0.9851125	0.9652932	0.9770874	uncharacterized protein	VLDQQPGSLECGYYVIRFVYDIFTKHRDSQDLTT DYSST

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
29	sp_049910_uxku	0.976905	0.98905	0.960653	ER lumen protein retaining receptor family protein; FUNCTIONS IN: ER retention sequence binding, receptor activity; INVOLVED IN: protein retention in ER lumen, protein transport; LOCATED IN: integral to membrane; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: ER lumen protein retaining receptor (InterPro:IPR000133); BEST Arabidopsis thaliana protein match is: ER lumen protein retaining receptor family protein (TAIR:AT4G38790.1); Has 889 Blast hits to 889 proteins in 227 species: Archae - 0; Bacteria - 0; Metazoa - 314; Fungi - 182; Plants - 232; Viruses - 0; Other Eukaryotes - 161 (source: NCBI BLink).	MGDKRATPTAVKQLSRWIRSRSVRQQVGMASM AAFILLILMKVFVKNYTYFYVAAQLAHSIGILVLIYK LTITKTCSGLSLKTQELTSIFLIMRIICHFMYIRDIYI VLDSITLGSTLWVIYMIRYKLKNTYLQDLDNMPFY YVLVPAFVIAFVGNPGVGISGLFYMYSKVIEAVSV LPQLRLIQNAKVVEPFTAHYVFALGVARFFSCAN WIILLIDTKGTMLTSIGSGHLYLLIILLAEVVQTFILA DFCYYYVKSVVEGQSMMRLPSMV
30	sp_049940_mmut	0.96948	0.933489	0.96907	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT5G53990.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAASNKEQSKLHIAMFPWFAYGHINPFIQLSNKL SSHGIQISFFSIPGNIDRIKSSLNLSPPNQLIPLTIP PTEGLSPNFDSSSEVTPQTAQLLTLALDQMQPQ VKALFPHPQPQVILFDFAYHWLPSVASELGIKAV HFNTFPAVINSYLTVPSRMTDPNKPPTFEDLKNP PQGYPKTSTASVKTFEAQDYLFLFKSFDGGPCH FEKILAFTNSCDAILYRTCNEIEGPFIDYFKTQINKP LLLAGPSVPLPPSGELDEKWEMWLGKFPEKSVI YCSFGSETYLNDAQIQELTLGLELTGLPFILVLNF GTSNSTDAHNKLEASLPEGFRERIKDRGVLHTG WVQQQNILAHRSIGCFLTHAGFSSVIEGIVNDCQ LAFLPLKADQFMIAKLFSGDLKAGVEVNRRDEDG SFAKEDIFEAIKTIMVDTDKEPSRSIRENHSNWRK FLMNKEIEASYIANLAHELKALVQKA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
31	sp_051630_ssst	0.954573	0.922223	0.977075	Major facilitator superfamily protein; FUNCTIONS IN: carbohydrate transmembrane transporter activity, sugar:hydrogen symporter activity; INVOLVED IN: transport, transmembrane transport; LOCATED IN: integral to membrane, membrane; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Major facilitator superfamily (InterPro:IPR020846), General substrate transporter (InterPro:IPR005828), Sugar/inositol transporter (InterPro:IPR003663), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT2G48020.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MKEEIKVQLLLNTDNNDKLVDEAGHHGLFPPLPE ANDTSLAVAAAAAVTPMLLFSVFIAVLGSFSFGFV VGYTSPVQAEIMKELDLSTAQFSLFGSLLTVGGA LGSVVCGKLTDYIGRRRVLQLTSLVFIFGWFAISV SQGAWLLDLGRWMLGFTGGISGFAIPVYVAEFT PKNLRGGFVILHVLMITVGVSAAFLIGLVTSWRTL ALIGIIPSLVQIVGIFFIPESPRWLMMMGKNKEFEA ALQCFRGGFVDITQEAADIRNSAEALDQIEKVSIL QLFHRKYAYAHIVGIGLSALGALVGLNGIIFYAGSI FESAGFSVKVGTVALASFQLPSVVLGVILMDKCG RRPLIMISAAGLFIGCSLTGLAFLFEDHHFLGGISP SIALIGILIYVAFYPIGVGGGTCIIISEIYPLNIKGSAG SIAAVVSSLSSWIVSYAFNFSMEWSSSGTFFVLA GVCVFTLIFVAKLVPETKGKTLEEVHKTMSDVWQ
32	sp_052880_qnfy	0.975383	0.979165	0.946745	XH/XS domain-containing protein; CONTAINS InterPro DOMAIN/s: Domain of unknown function XS (InterPro:IPR005380), Domain of unknown function XH (InterPro:IPR005379), Domain of unknown function, putative Zinc finger, XS/XH (InterPro:IPR005381); BEST Arabidopsis thaliana protein match is: XH/XS domain-containing protein (TAIR:AT1G15910.1); Has 52398 Blast hits to 32322 proteins in 2074 species: Archae - 615; Bacteria - 6626; Metazoa - 25839; Fungi - 3724; Plants - 1961; Viruses - 202; Other Eukaryotes - 13431 (source: NCBI BLink).	MENCNQEDELKTLRNLVVKLARELDYKNFTLMEL EEKNAKDLENWMKEKEKLLIERSAAISVLLSENV DKDLLLDEKSARVIALTEEKEQMLRETTKSMNTL KEESERQINEKSAVINTLKAQNADKERLLEEKEW LLDERLTCMNTLVVEKERLVNEKSAVIRTLMIQNA EKEMLNEEKSAAISKLMAEKEQGLEALQKELKKT ENMKLKNEKLQCDLESQKGELERLIKEKNKLMQE TNNTELEVVEYCSKPCKGELNVQPREPEDSRVQ TVEEINCSMPLQMQNPEDLRTEIVILRKDLEGKAE LLESMEMENRILMVKELKSNHELQVARKAATEGL LAMHRPRMNIGIKKMGEVNLKPFRDACSKRFRS GNWDEKFGNWEEKSVMLCSSWQREISDPGWQ PFKQETVNGKLTEVIDENDEKLRELKEEWGEEAY KAVKEALVEVNEYNASGRYPVSELWNHKQNRKA TLEEVIQYVIKQWRTNKNKKMAV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
33	sp_056760_rown	0.957653	0.955079	0.964373	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: nucleobase, nucleoside, nucleotide kinase activity, nucleotide kinase activity, ATP binding; INVOLVED IN: nucleobase, nucleoside, nucleotide and nucleic acid metabolic process, anaerobic respiration, nucleotide metabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Adenylate kinase (InterPro:IPR000850); BEST Arabidopsis thaliana protein match is: adenosine kinase (TAIR:AT2G37250.1); Has 9469 Blast hits to 9393 proteins in 2915 species: Archae - 95; Bacteria - 5951; Metazoa - 891; Fungi - 419; Plants - 430; Viruses - 0; Other Eukaryotes - 1683 (source: NCBI BLink).	MAVLSRISVTSPSLRHLLHLSPPLFQLATIRTYRIT AAATAAQLQFDYDDDFGRTPPSPTPATAVDTDV SRPFRGVHWVFIGNPHAKKHIYADMIAKLLQVPRI SISTLLRQDLNPHSPLYKQIAEAVNCGKLVPESVI FGLLLKRLEDGHDRGEDGFILDCVPRTRLQAEILD QLVDIDLVVNFRCSENHFQATGSPATSLDRTKSS FSKLNKTAEDSQTKVSAFYREQMQYLEDYYQSQ KKLIDFQVGNASGETWKGLLAALHLQHTNAVHS SRTLTMGFRVP
34	sp_060610_zcnc	0.976479	0.995916	0.966373	Phosphoglycerate mutase family protein; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histidine phosphatase superfamily, clade-1 (InterPro:IPR013078); Has 1391 Blast hits to 1391 proteins in 390 species: Archae - 1; Bacteria - 792; Metazoa - 0; Fungi - 0; Plants - 44; Viruses - 0; Other Eukaryotes - 554 (source: NCBI BLink).	MGHNKGWEETASMFTGSVIELKTCNAALLEGTG KSWEEMILSFLLPFALVPLPKFTSGKAKMGVHAN SFMAPG
35	sp_061210_ojgo	0.979939	0.982438	0.996827	YELLOW-LEAF-SPECIFIC GENE 8 (YLS8); CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), mRNA splicing factor, thioredoxin-like U5 snRNP (InterPro:IPR004123), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: mRNA splicing factor, thioredoxin-like U5 snRNP (TAIR:AT3G24730.1); Has 564 Blast hits to 564 proteins in 220 species: Archae - 0; Bacteria - 0; Metazoa - 221; Fungi - 146; Plants - 112; Viruses - 0; Other Eukaryotes - 85 (source: NCBI BLink).	MSYLLPHLHSGWAVDQAILAEEERLVIIRFGHDW DETCMQMDEVLASVAETIKNFAVIYLVDITEVPDF NTMYELYDPSTIMFFFRNKHIMIDLGTGNNNKINW ALKDKQEFIDIVETVYRGARKGRGLVIAPKDYSTK YRY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
36	sp_064650_ufdn	0.922751	0.967018	0.945118	receptor-like kinase in in flowers 3 (RKF3); FUNCTIONS IN: receptor signaling protein serine/threonine kinase activity, kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT1G11050.1); Has 112833 Blast hits to 111726 proteins in 4356 species: Archae - 97; Bacteria - 12546; Metazoa - 41926; Fungi - 9138; Plants - 32648; Viruses - 383; Other Eukaryotes - 16095 (source: NCBI BLink).	MCHLHLLLFLSLQLLLLSSAQNDTVPSTSSSSSP CPLNFTTLSTLVRDADVKPSNDVGANCGYIAEGL RLVLSQYLQSTGSFYPPVNTSESCWTSYDSVVD SFLPGFKIRSSCGFRTQWISEGCMNISTRSEFER LVPQSALDDAVSNCNQSLENGSPCASCLTSLTTL QASYLTSNSPFNGNISDCTNLPLIYAGALINQYGP SDPGTAKCLFSLDFSSSQNFNNRKKKILLISLITVV GFIGICSIFVGGYFFSRLKKNRAKMAEKKRRINRV GSGAGPNLSSIGGSTHLMKFGYDEIKRATKNFHR DNLIGKGGYGNVYKGLLPDNTTEIAVKRFKNCSA AGDASFTHEVEVIASIRHVNLVELRGYCVATTNLE GHQRIIVCDLMHNGSLHDHLYGSTKLSWPFNPE GFSHLSTRVAGTMGYVAPEYALYGQLTERSDVF SFGVVLLELLSGKKAILSDAGNPSLSPSLVTDWA WSLVRKGHGLDVIEEGMPELGPESVMEKYVLTA VLCSHPQLYARPTMDQVVKILETENMQVPAIPER PISFNAELEDIERSASSSGSGQMSTSGGFQSYYS MGHVDFGSHNGLEEVEDN	
37	sp_066190_ewgc	0.972228	0.947988	0.983387	COBRA (COB); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to salt stress, multidimensional cell growth, cellulose microfibril organization; LOCATED IN: in 6 components; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl-phosphatidyl inositol-anchored, plant (InterPro:IPR006918), COBRA-like (InterPro:IPR017391); BEST Arabidopsis thaliana protein match is: COBRA-like protein 1 precursor (TAIR:AT3G02210.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MESCCSCRSISKLTSFTLFIAFLLSCSFFTSSEAY DALDPNGNITIKWDVINWTPDGYLAVVTMFNFQQ YRHIQTPGWTLGWTWAKKDVIWSMVGAQTTEQ GDCSKFKGNIPHCCKKDPTVVDLLPGTFPNQQIA NCCRGGVLTAWGQDTANAVGAFQISVGQAGTS NKTVKLPRNFTLKAPGPGYTCGPAKVVRPTQFT TADKRRTTQALMTWNVTCTYSQFLAQKTPTCCV SLSSFYNDTIVPCPTCSCGCQNNITQPGSCVDGD SPYLASAISAGGPKASNTPLVQCTSHMCPVRIHW HVKVNYKDYWRAKVTITNFNYRMNYSNWNLVVQ HPNFDNLTQSFSFMYKGLTPYDGLNDTAMLWGV KFYNDFLNQAGPFGNVQSELLFQKEKSSFSFEK GFAFPRRVYFNGDNCVMPPPDAYPWLPNASSR PIISLLQPVMTILAALAFLLAYA	
38	sp_070160_ugjc	0.952031	0.963398	0.977097	uncharacterized protein	MLSQFSSSMEGAIKGRRLKTSMMMIGLGTRQVV LPKYVMLMLQKTEGLRQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
39	sp_074730_rlxn	0.993446	0.944573	0.8502813	mevalonate kinase (MK); FUNCTIONS IN: mevalonate kinase activity; INVOLVED IN: isoprenoid biosynthetic process, metabolic process, phosphorylation; LOCATED IN: cytoplasm; EXPRESSED IN: 30 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Mevalonate kinase (InterPro:IPR006205), GHMP kinase, ATP-binding, conserved site (InterPro:IPR006203), Mevalonate/galactokinase (InterPro:IPR006206), Ribosomal protein S5 domain 2-type fold (InterPro:IPR020568), GHMP kinase (InterPro:IPR006204), Ribosomal protein S5 domain 2-type fold, subgroup (InterPro:IPR014721), GHMP kinase, C-terminal (InterPro:IPR013750).	MAVTTRAPGKILAGEHAVVHGSTAVAASINLCTS VSILFSPGDMLKLELKDMELEFSWPIGKIKEVIPD FDTPSPSTPLPCSSETLKSIASLYEEQNIPEAKITL ASGVSAFLWLYTATQGYKPATVTVTSELPLGAGL GSSAAFCVALSGALLALSECCELDYSQQGWLTL RNSELDLVNKWAFEGEKIIHGRPSGIDNTVSTYG NMIQFKSGEITRLKSNTPLKMLITNTKVGRNTKAL VAGVSERTLRYPEAMTCVFKAVDSISKELSNIIQT SASNDIAVTEKEDKVEELMEMNQGLLQCMGVSH ASIETVLRTTLKYKLRSKLTGAGGGGCVLTLLPTL LSSIVVDKVVAELESCGFQCFTAGIGGNGLEVCF SDNAQI
40	sp_074810_agcn	0.9772032	0.9353386	0.9603715	Arabidopsis protein of unknown function (DUF241); LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF241, plant (InterPro:IPR004320); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G01590.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAASSHPVRSISLFARSHPTSLKIEETINKLKSRDI YQSDSTSCYTSGSDSIQLGLLNLVELYARVQELV SSLSSQHGKLVEEALESSVSLLDTCSIARDLISGF KEQVCDLQSALRRKGSSGTEKEINIYMSFRKKVS KEVSRCLKVLKQAEAKYGITAPLLDNDPSLVNLN RLLRETYTATIVVFKSLFLFLSTPTKPTKWSLISKL VTKNSISKEAVTQLNEVSGVDSAIQNESKVDVML LRGKIQALDSSLQILEGEVGCLFRCLLQNRVSLLN ILTQSI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
41	sp_076690_ejcm	0.961713	0.923419	0.953834	cellulose synthase like G1 (CSLG1); FUNCTIONS IN: cellulose synthase activity, transferase activity, transferring glycosyl groups, transferase activity; INVOLVED IN: cellulose biosynthetic process, polysaccharide biosynthetic process; LOCATED IN: membrane; EXPRESSED IN: hypocotyl, sepal, male gametophyte, flower, carpel; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Cellulose synthase (InterPro:IPR005150); BEST Arabidopsis thaliana protein match is: cellulose synthase like G3 (TAIR:AT4G23990.1); Has 3059 Blast hits to 2371 proteins in 475 species: Archae - 2; Bacteria - 903; Metazoa - 2; Fungi - 10; Plants - 2111; Viruses - 0; Other Eukaryotes - 31 (source: NCBI BLink).	MATSHIRNVQLTRAIVNRLHIFLHSVAIFSLFYYRF TSFFNSDISILAYSLLTTAELFLTFLWAFTQAFRW RPVMREVSGYESIKPEQLPGLDVFIVTADPTKEP VLEVMNSVISSMALDYPVDRLAVYLSDDGGSPLS KEAIKKAYEFAKLWIPFCNKYNVKTRCPQAFFSPL ADGERLDWNSEFMADQLELQTKYEAFRNYVEKE SGDNTKCTAVHDRPPCVEIIHDNKQNGESDVKM PLLVYVAREKRPGRPHRFKAGALNALLRVSSLMS NAPYLLVLDCDMYCHDPTSARQSMCFHLDTNMA SSLAYVQYPQIFYNVSKNDIYDGQARSAHMTKW KGMDGLRGPVLNGTGYYLKRKALFGKPNNEDEY LNSQPEKAFGSSTKLIAALRENSKQNLAIKELTED ELYQEARNLATCTYEANTLWGSEVGYSYECLLE STFTGYMLHCRGWKSVYLYPKRPCFLGCTTIDM KDATVQLIKWTSSLLGIALSKSSPLTLAMSSMSIL QSMCYAYITFTGLFAAPLVIYGVVLPISLLKGFPIF PKVSDPWILFFVLIFVSSHLQHLYEVLESDKSATQ WWNEVRIWMMKSVTAGLFGLTEAIMKKIGVQTA TFRLTNKVVEKEKMDKYEKERFDFSGAAMLMVP LNILVVLNMVSFIGGLMRVIINNSYDQMFAQLFLS FFVLLLSYPVVKGWL	
42	sp_079260_kmfz	0.9881837	0.9444412	0.9774031	Regulator of chromosome condensation (RCC1) family protein; CONTAINS InterPro DOMAIN/s: Regulator of chromosome condensation/beta-lactamase-inhibitor protein II (InterPro:IPR009091), Regulator of chromosome condensation, RCC1 (InterPro:IPR000408); BEST Arabidopsis thaliana protein match is: Regulator of chromosome condensation (RCC1) family protein (TAIR:AT3G53830.1); Has 20178 Blast hits to 6212 proteins in 476 species: Archae - 78; Bacteria - 2696; Metazoa - 7040; Fungi - 1181; Plants - 2975; Viruses - 2; Other Eukaryotes - 6206 (source: NCBI BLink).	MNGGEGGGGGESEVQMEIEMREEERKKRVYMW GYLPGALPQRTPLQSPTLVKLPPSSGNFWKDVC GGGCGFAMAISDTGKLITWGSTDDLGQSYVTSG KHGETPEAFPLPTEAPIVRAAAGWAHCVSATDR GEVYTWGWKECVPSGKVFGDSSSGGSSEKDAL EKQSSLFPEQVSPRGSKNANGIGPGAVLENKGA GEETTKRRRVSSSKQTVESSTASDETLTALPCLV TLNPGVRISTVAAGGRHTLVLSASDVGQVWGW GYGGEGQLGLGSRIRMVSSPHPVPCLESSYRKD RHSGFSQGGLNSDGHGFRVPGTYIKGIACGGRH SAAITDAGAILAFGWGLYGQVNRFDREKSRVDV QDEAHSSQRSDDYFLPTLCASLCGQGATDDVLS PTCVSSLLGIRIEAVAAGLWHTVCISAEGDVYAFG GNQFGQLGTGSDQAETLPRLLDAPSIESVNATIV SCGARHSSIITGDGKVFSWGWNKYGQLGLGDAI DRNIPSQVKLDGSVAAKNVACGWWHTLLAESPT CD	
43	sp_079290_hwwi	0.9965356	0.963583	0.9911348	uncharacterized protein	MAASTTSLPSTLTLPLFILILLLFNYTSSGSSMAIG SSRNLQAASVHKQTDNTQKNQIPNCSEIVAKSHC SSNPKCRWCRSDYLDSTCSTTSDAWRFPSQIFI CD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
44	sp_079580_zhwh	0.957886	0.981127	0.96935	Major facilitator superfamily protein; FUNCTIONS IN: transporter activity; INVOLVED IN: oligopeptide transport; LOCATED IN: membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oligopeptide transporter (InterPro:IPR000109), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT3G16180.1); Has 5212 Blast hits to 5001 proteins in 870 species: Archae - 0; Bacteria - 1843; Metazoa - 575; Fungi - 389; Plants - 2143; Viruses - 0; Other Eukaryotes - 262 (source: NCBI BLink).	MREPLLGTSNGGLKTLPFIIANEAFEKAASFGLLP NMILYLMRGYGMELATGSNIIFFWNAATNFTPVIG AFLADSYVGRFWMIGFGSIFSFLGMIFLWLTAIIP GATPCDDSGSECKSWNVFQLTLLVSAFGLMSIG AGGIRSASLAFGADQLLKKDNFSDAGILERFLNW YYSACCGAVVVGVLFVVYIQDHFGWKVGFGVSV VLMLLSALSFFSASTLYIKFKAKAGSLIAFAQVLMV SWKKRNLNFSEDELYYQSKGSRLLIPSDKLRFLN KACIVGNAEEELTPDGVAIDPWNLCTVNQVEEVK ALIKVIPMWSTGIMMSVTTTQFTFPVLQAGSMDR HITSNFQVPAGSFGTFTVISIILWIAIYDRIVIPLATK ITGKPGQLSVRKRMGVGLILSFMSMVVCAIIEGVR RETAIKEGFSDDPKAVTDMSAYWLLFHHWISGM AEALFLVGQNEFFYSEFPKSMSSVATTISGLSTS VASLFASLIMSSINSITSKGGKESWVSSNINRAHF DYYYWLLAGLSFLNFLYYLLCSRAYGPPSYDQDR FEDNEEIIEG
45	sp_080440_xyfo	0.951073	0.990295	0.950668	FUSED (FU); CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Armadillo-like helical (InterPro:IPR011989), Armadillo (InterPro:IPR000225), 26S proteasome non-ATPase regulatory subunit 5 (InterPro:IPR019538), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: serine/threonine protein kinase 2 (TAIR:AT3G08720.2); Has 138334 Blast hits to 135590 proteins in 4834 species: Archae - 164; Bacteria - 15793; Metazoa - 51363; Fungi - 13983; Plants - 33508; Viruses - 544; Other Eukaryotes - 22979 (source: NCBI BLink).	MGIASNYHMMEPVGEGSFGKVYKGRRKCTGQV VAMKFILKNGKTEKDLCGLRQEIEILKKLKHENIIE MLDSFETAEEFCVVTEFAQGELSKILEDHEVLPE AQVQAIAKQLVSALHYLHSNRIVHRDMKPQNILIC SGSYVKLCDFGFARAMSTKTVVLHSFKGTPLYM APELVREQPYNHTVDLWSLGVILYELFVGLPPFS ADSFYSLFRLIVKDPVKYPENMSLEFKTFLKGLLN KIPQQRLTWPTLLEHPFIIGSPHLLKDLPSTAKAA LKGKTPVIPLHPDTQNFESGQALDRFENNSSTLE GAKTIGQEMCALELILEPVMMLSKGTHNCCCHSL VQADIRDCDGLPDIAGGSDKVVETITRAFLNSEPV QDAMWYCLCQCKDEVLLATIQLLHRCCMHSGIL SNVLCGFPRSLPAATAISGEGKTFVSEIFSIISQCG SSLNKDLPTSEAIGSKFEYTNISGLALHSCLLLATI AQSFKSAVGNSAALILTTSSEIQQSRLSSLACLFS FNDGAVTSLQPHCASAMLAFASIILIESGASVNSSI MLERVHVNCINDLISGAHLKLLKSWCGPGGGLN GVREIIDAAIDFLAFPFVATENSLGYPSVTASMYS GIPLNVSSPAANLYMEEGDNKKEIKENMGKYVEIL IENLYKYIERADILQRLKELLCSDDCAIRAKVCNAV GNMCLHSSYFYHLLGKYGIISLLIDRCTDSDLQTR NFACFAVGNAAFHDDLLYEELRRSIPQLTNLLIST EKDDIKVNAAGTLYSLVRHSNKLGDEIMSNGAIQ ALLKVVSDECSKLALNSLKMDAIDASPLKNALCAL AKMCTCYPPCRNVIFSSESFQVIAQLRQSPAPKIA KFARDICQATGRELEAYT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
46	sp_084440_pkxa	0.928427	0.947385	0.973345	cytochrome P450, family 71, subfamily A, polypeptide 21 (CYP71A21); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: oxidation reduction; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, conserved site (InterPro:IPR017972), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 71, subfamily A, polypeptide 22 (TAIR:AT3G48310.1); Has 32159 Blast hits to 31925 proteins in 1622 species: Archae - 46; Bacteria - 2978; Metazoa - 11709; Fungi - 6921; Plants - 9489; Viruses - 3; Other Eukaryotes - 1013 (source: NCBI BLink).	MSSLFLQLLNLFFHPLSLLVIFFVIYFYKSLSAKPA TAKNLPPSPPKLPIIGNLHQLGLLPHRSLQSLSQR YGKLLFIYIGSKPSLVVSSADAAQEIMKTHDMIFS NRPKSLAASRLFYGGKDVAFSPYGEYWRQMRCI CVTQLLSNTKVQSFRKIREEEAALMVESIERCGS SVVNLREILMTFTNDVICRASFGMKYDHEGGFNL KKLLDKVMELLGDFCVGDFIPWLSWIDRVSGLEG RLEKVAKFLDAFLEKVIQEHQIRQSTKIHHNELDK VKNVQNFVDILLEVQSENREALPLDSIKAVILQPS HNPTLKFWSIAPGIKSLPRLQSDRYALK
47	sp_085210_mwkt	0.870138	0.945105	0.980164	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, kinase activity, ATP binding; INVOLVED IN: transmembrane receptor protein tyrosine kinase signaling pathway, protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: shoot apex, embryo, leaf whorl, sepal, flower; EXPRESSED DURING: petal differentiation and expansion stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat protein kinase family protein (TAIR:AT5G35390.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MTVIFILLTKYRSPKYGTGKKITARDIETAPHEKAK LKQSNGHDTYPNNQARALERATSYKKSDNGKLC FLRNDRQKFEMQDLLKASAEVLGSGSFGSSYKA AIMGGPTFVVKRFRQMNNVGKEEFQELIRKLGRL RHPNLLPLVAFFHKRDEKLIISDFVENGSLASHLH GRHTPDQPGLRWSMRLKIIKGVARGLAYLYRELP ELSLPHGHLKSSNVLLDEKFEPLLSDYAFVPLINK EHGEQVMAAYKSPEFTQTNAVTKKTDVWCLGILI LEILTGKFPANYLKHGKGGNDELAIWVDSVVKEE WSSEMFDMEMRASRSCKGEMLKLLKIGMWCCE WNEDKRWSLRDAVEKIEELKEMDDDNLADFPSG RSYSARSFSSDSSYSRIASITTEQVV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
48	sp_085340_meek	0.967688	0.975143	0.9622	cytochrome P450, family 72, subfamily A, polypeptide 15 (CYP72A15); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: response to cyclopentenone; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 72, subfamily A, polypeptide 13 (TAIR:AT3G14660.1); Has 31875 Blast hits to 31741 proteins in 1631 species: Archae - 65; Bacteria - 4943; Metazoa - 11298; Fungi - 6149; Plants - 8001; Viruses - 3; Other Eukaryotes - 1416 (source: NCBI BLink).	MIEIGYIVKWVICLVIVRWVWKIVNWVWFTPKRL EKFLRKQGLDGNSYRFLLGDLKDMSKMRKEARQ KPIPFTHDFFHRILPFHNHHFNKYGESFFSWMGPI PVVNVAEQEQVKGVFTRIKEFQKAKLNPLVALLV PGLVSAEGDKWVKHRKLINPAFHMEKLKLMHPA FGASVLDMVNKWEKIVSKTGSSEVDVWPFVSSL TADAISRAAFGSSYDEGRKIFELVLEQTEITLRLLQ SVYIPGWMYVPTKTNRRMKTVNSEIQNLLTGIIVK RKKAMEAGEAAKDDLLGILLESNYKDTENVLSNK KKLSMTFQELIDECKLFYLAGQESTSVLLAWTMIL LGKHTEWQARAREEVVATFGKNEPDFEGLNHLK IVTMILNEVLRLYPPVCTITRKNFNHDVQLGNLTV PRGAMVTMSAYRIQRDPKIWGDDAKEFNPQRFS EGVAKATKGNIAFFPFGWGPRICIGQNFALIEAK MAVSMVLQRFSFELSPSYTHAPTTILTLQPQQGA HLIIHKLRD
49	sp_085760_qcir	0.935788	0.985641	0.938626	uncharacterized protein	MEDDDDLDQDFLNMCGNIANNDEFMSNDELLYE FQTAADIVAEPPVIPTVINVLRVVFFGLIYVMKEGI GVIYLFILFGSIF
50	sp_087690_iwxz	0.976794	0.958753	0.981434	lactoylglutathione lyase family protein / glyoxalase I family protein; FUNCTIONS IN: calmodulin binding, lactoylglutathione lyase activity; INVOLVED IN: response to cadmium ion, carbohydrate metabolic process; LOCATED IN: chloroplast; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Glyoxalase I (InterPro:IPR004361), Glyoxalase I, conserved site (InterPro:IPR018146), Glyoxalase/bleomycin resistance protein/dioxygenase (InterPro:IPR004360); BEST Arabidopsis thaliana protein match is: Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily protein (TAIR:AT1G67280.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MESASETKESASNNPGLHSTPDDATKGYIMQQT MFRIKDPKLSLDFYSRVLGMSLLKRLDSPTMKFSI YFLGYEDVASAPSVPVDRTVWTLSKKATIELTHY WGTESDPEFKGYHNGNSEPRGFGHIGITVDDTC KACERFEQLGVEFIKKPDDGKMKGITFIKDPDGY WIEIFDLKTIRSVAAGAS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
51	sp_088530_wpzu	0.928893	0.934266	0.96904	polygalacturonase inhibiting protein 1 (PGIP1); FUNCTIONS IN: polygalacturonase inhibitor activity; INVOLVED IN: signal transduction, defense response; LOCATED IN: cell wall, plant-type cell wall; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: polygalacturonase inhibiting protein 2 (TAIR:AT5G06870.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MPSNRDEMSLFSVFILILLLSSTISADRCHPDDKK ALFRIKKAFNNAYIFASWTHDTDCCEWYLVKCDD TTHRIYSLAVSSDDKVSGHIPDAVGDLPYLESLTF LRLPKLTGPIPQSILKLKNLNFLWLSHNNLTGPIPD FLGYLPKLTYINLSVNKLTGHIPSSLSKLHALGALF LERNYLSGPIPESFQGFKNNPDFYLKLAKNQLSG PIPKSLGGINFTLLDISRNMLTGDASVFFGRNKGL QTVDISRNKFSFDLTKVGLPKTLNTLELSHNMIYG NLPPTLAKLPMLQRFNVSYNRLCGKIPTGGDLNR FDKYSYAHNKCLCGSPLPPCK
52	sp_089620_qnin	0.933658	0.927806	0.981159	non race-specific disease resistance 1 (NDR1); CONTAINS InterPro DOMAIN/s: Late embryogenesis abundant protein, group 2 (InterPro:IPR004864); BEST Arabidopsis thaliana protein match is: Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family (TAIR:AT3G20590.1); Has 471 Blast hits to 471 proteins in 25 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 471; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGVEEGGCCRCCFSFIFTSGLTALFLWLNLKTTN PTCSIENFDVFTLNKTANTTFRNNHTIRYDLKLNN KENKDKGIYYDTLNLTFYYKPNLSFVTLGNATFNP FYQGHGKNTHRISVIDARGVKWENATATMAEFR VEMNTSVRFKIIFWKTKRHRLDLHADLLVNEQGIL VKKKKKKGIKFKSGVWNSKVCNFYLLGILGIFTYIL LW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
53	sp_090460_pfyf	0.926438	0.922207	0.958825	NB-ARC domain-containing disease resistance protein; FUNCTIONS IN: protein binding; INVOLVED IN: apoptosis, defense response; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing protein (InterPro:IPR015766), NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: LRR and NB-ARC domains-containing disease resistance protein (TAIR:AT3G14460.1); Has 25652 Blast hits to 19792 proteins in 681 species: Archae - 38; Bacteria - 1467; Metazoa - 4997; Fungi - 304; Plants - 18213; Viruses - 0; Other Eukaryotes - 633 (source: NCBI BLink).	MECVVLPLLTEAAKPMIEVLVSRVTAEIHSFRGFE KDLSKLQRKIQIISNIVLHMTPTNTNTNSRDPVEL WVRNVVKVIHNVENLMDVLSYEKQHQVVMKRRL RTSFNGKTTKLLFWPVKRVYFRLKIGKWISRINTN IEEALKEAETISQLRGLTHSPSLLPAATNIITSAEN RLQQLRRNVDMSQIVGRDSDVDTVVDKLCNPSH DFYNVSSNRLSLIAIVGIGGIGKTVVAKKVFEDKR VMDNFTERIWVVAPAVFSSEGILNKMLEYIYMSN NQEMWDATRSSLECIGGSHESMILVTTRATVVA ETMHAFVVHSLPELSEDDSWSLFNRVAFAKMSP EYISILETTGKKVVQKCKGVPLAIKSIGGMLQKRS HPYEWEKIERSEIWDLLEGDTGILPSLMLSYKNL SSSFLQKCFSFCATLPKGSHMARSVSKHELLLLE ADNQVTAISDVRHLAIYEHKELKTDCWDKEKVAI ENLRTLAIYKNSPGGLLKHVKYLRVLDLEDMGLK EVPECIAGLKCLKFLGLVNNPIKEFPEFITKLFNLQ SLHFNECDQIKEVPGGLSNLVNLRHLDIPGLIKLP RGVIGQLTCLQTLPILGVREGGFQISELGGLPHIS DKLEIQGLELIKSKEDAKSASLSEKSRVKKLRLFW NTDNDDDWVKHEEVLDGLKPNRNLRLLDIQGYG GENFPSFWLMRMDMIKIIELTDCRRCRKLPTLGH LHSLEKLKIKNMESLECIGVELYCYTYTVGASSIS SSSSSSNSEVKYFPSLKQLQLLGLEKLTEWIEPS SGEWNIFPLLEELQMEDCPELKTTLYKFHSLEHL TVKNITSGLLLSKILSNKGSVTCLKIVQIEGITEITD LPEELAEYCTSLKTLRIHRCQNLCRLPESLSKLISL EVLQIKVCPNLVSLPSLSSLETLRELEFYNVGLTS VPEGLQSCKSLERLWVDCCPKVETLPDLTGLSH LCDLNIRRMTKMINQPPDWLYRLPCLTKLTIGGYL GRNEFPDMSFLWKITSLEALNLCGWGKLQSLPE QIQHLTTLKRLGIYDFKNLEYLPDWLGNLSSLDAL STTRCENLRKLPTEDAMKRLTKLSYLTILECPLLE
54	sp_090690_mtjc	0.964945	0.986844	0.942604	uncharacterized protein	MARRTREGPKQGKGKHEEEGREEGEREAGLGG STPRLRTPKGGGEGGGLQRRYGEGRERKGNAN REGMGVSRGRKWMVSDGGGSPGIRAEVGFDD DGCGVVS
55	sp_095820_trmg	0.962008	0.932904	0.938048	uncharacterized protein	MKTELGIDTSSEQPSCLYSSVQKSSSNYGVCRTL FSNIGR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
56	sp_096930_munp	0.965654	0.931363	0.975462	uncharacterized protein	MGELPRFIVLGRPTGGVTYLRYVVERGAQHRRL HLSSPEVISERNFTKLEVVRAPPTIGRDMVHIRSC YNNKFLVLESEGSNVFVAEADEPEVDRSIWSCTL FRVTYVNNRLSLLLQSSNAQHSVQVLSNSSLGIV RNSPFAYWFDVTNMSTLAILPEYIAFKGDNNRYL NSRHIRTQNRLKYESTNLLDNGVHHQTQTLRNG NVRFRSIHFGRFWRKSNPEWIFGDSDDRTGSDI NTVFRPIKLDGNKISLLNLGNNRYCMRRVQGGQP KLVDTLTASSSTVGPLTVMEVEEAVRQREIYNIHY NMDESRVYNVAPFESGNVTQDTYNLTPLPVVKR VGFKYTQVTRETFNSSHSWKVAYNSKIKVTVIPK FVEGEATLEASYDGKVEWSELNETIITQTKDLFIP VPPMTKSIVTLILGRGSVDVPFSYEQRDILPNGRE VHKHMEDGLFSGLNTEVFTHTVREISLSTLSESE LEEARRLSVVDEEAAAPALSSKL
57	sp_097340_nzpp	0.978407	0.953333	0.975906	Class-II DAHP synthetase family protein; FUNCTIONS IN: 3-deoxy-7-phosphoheptulonate synthase activity; INVOLVED IN: aromatic amino acid family biosynthetic process; LOCATED IN: membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: DAHP synthetase, class II (InterPro:IPR002480); BEST Arabidopsis thaliana protein match is: 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 1 (TAIR:AT4G39980.1); Has 3811 Blast hits to 3796 proteins in 685 species: Archae - 0; Bacteria - 1237; Metazoa - 0; Fungi - 110; Plants - 171; Viruses - 0; Other Eukaryotes - 2293 (source: NCBI BLink).	MALSGNTILPTKTTFLPSTTTAKPTTFSLLPLKPR HISAVQSSDKNHPITSSKPTIPSAAAASTTAISSIA KKSSSEQQWTLESWRSKKALQLPEYPDQKELD AVLKTIEDFPPIVFAGEARILEERLTDAALGNAFVL MGGDCAESFKEFNANNIRDTFRVLLQMGAVLMY GGQMPVVKVGRMAGQFAKPRSDGFEVRDGVKL PSYRGDNINGDAFDEKSRIPDPQRMIRAYCQSAA TLNLLRAFATGGYAAMQRVVQWNLGFAERSEQ ADKYQELAHRVDEALGFMAAAGLTLDHPLMQST EFWTSHECLLLPYEQSLTRKDSTSGRYYDCSAH MVWVGERTRQLDGAHIEFLRGIANPLGIKVSDKM DPNELIKLIDILNAENKPGRITIITRMGAENMRVKL PHLIRAVRRAGQIVTWVSDPMHGNTIKAPCGLKT RPFDAIRSEVKAFFDVHEQEGSYPGGVHLEMTG QNVTECIGGGQTVTFDDLSSRYHTHCDPRLNAS QALELAFIIADRLRNRRMKSQASLTS
58	sp_098440_zepn	0.955385	0.985553	0.949899	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: nucleotide binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: pentatricopeptide (PPR) repeat-containing protein (TAIR:AT5G27300.1).	MPRYKDEPPTVRVYTVCDESKYLIVRNVAALGCG DELRQLFAAYGDVEECKPMDAEDCEAFTDVYWI KFHMISNARFAKRKLDESVFLGNRLQVSYATEYE TLSDTKEKLEGRRKEVLGRLNSLKSGRYRGDEM KSKGVETHPLSHGAGVSSQVSQQLDLKQGWYA KEMNQYSEVKHSSITRVSSDQDYFPSQSMNETV KRVREKLDKIESSSSSNTQSGSASKKIRVDNRRR I

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
59	sp_099260_aaoe	0.974225	0.921695	0.961889	Polyketide cyclase/dehydrase and lipid transport superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to biotic stimulus, defense response; LOCATED IN: cellular_component unknown; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Bet v I allergen (InterPro:IPR000916); BEST Arabidopsis thaliana protein match is: Polyketide cyclase/dehydrase and lipid transport superfamily protein (TAIR:AT5G28000.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSSLKRKLEGEIEIRAAAGDVYHDIYKSRPHHLSH ASPHFVQSCDLHEGELGKPGSIIVWKYKLGGKAL TMKCVIEEIHEEKKLVRQRVVEGDSLEEYKNLVG TCHVIAKDSETSIVRWMLEYEKMHAGIPEPSALL DACLEAAKDADDHHHGIKK
60	sp_102570_knxq	0.936867	0.956269	0.984536	Aluminium induced protein with YGL and LRDR motifs; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cytosol, nucleus, plasma membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: Aluminium induced protein with YGL and LRDR motifs (TAIR:AT5G43830.1); Has 1324 Blast hits to 1324 proteins in 483 species: Archae - 0; Bacteria - 791; Metazoa - 46; Fungi - 0; Plants - 333; Viruses - 0; Other Eukaryotes - 154 (source: NCBI BLink).	MLAVFNKEVAKCPEELKNAESLGNVKEDALIQQF SSVHPNALSVHLGDSGFIAFSSDKQNPLLPRLFA VVDDIFCLFHGHIENIPSLKQLYGLTKTANEVSIIIE AYRSLRDRGSSPADHALRNIEGKFAFVIYDSSIKT TFFSVDTDGSVPLFWGTDSEDHLVLSDIPEVMQ NACGRSFAPFPKGFFFTSAGGLKSYEHPRKVLK AVPWVDSSGQACGSTFMVTEESNKENKQLGIKS GMPRVGSDADWSKKI
61	sp_104090_cxdg	0.970797	0.995919	0.991434	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G02370.2); Has 122 Blast hits to 122 proteins in 54 species: Archae - 0; Bacteria - 0; Metazoa - 74; Fungi - 4; Plants - 41; Viruses - 0; Other Eukaryotes - 3 (source: NCBI BLink).	MGGRREAEELSNSLGGLSDSEYSSQDSADDEW CSEAGSETKLQCRSVVLKACWKDETGMAEVMN HKGNLWKTIGINRNGKMYLSIEETLFLIQIGALLLV DDKDTTFTLSEVYPKVAEEKSGCSWEGFKIYQHL KSLGYIVGRVGIPWSQKSTKNCTMVHEDTSETD NVTQIIFSRTSIVQPLNNLSLNEANPLSEDELPNKF KKTFPEESKFVVHLPVYEVYLPNSQFKKTCPGDP DFVVYPTSGHPPSVFDIKELERQSNGLPIKVGHV DEGRVSLFSFDKVELPVLA
62	sp_107890_mqkw	0.945801	0.945673	0.988099	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G22758.1); Has 131 Blast hits to 131 proteins in 17 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 131; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MSPENTVGHLIKAVLEVFEKERRRPLLNENHPSC FELHYSPFSLQNIYSCQIKIQKGEFYVSRRRSIKF QSTESDADT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
63	sp_107930_fsnu	0.944987	0.942288	0.938835	Zinc finger (CCCH-type) family protein; CONTAINS InterPro DOMAIN/s: Zinc finger, CCCH-type (InterPro:IPR000571); BEST Arabidopsis thaliana protein match is: Zinc finger C-x8-C-x5-C-x3-H type family protein (TAIR:AT2G35430.1); Has 13421 Blast hits to 5509 proteins in 634 species: Archae - 13; Bacteria - 2784; Metazoa - 4906; Fungi - 574; Plants - 3348; Viruses - 263; Other Eukaryotes - 1533 (source: NCBI BLink).	MIRQFGQLKMSIEIGITVKQPVTTYLHCSVWDQVA ELSLPIKRLGAPPRVAAMLRRXIGKMFYKTKLCCK FRAGVCPYVTNCNFAHGLEELRKPPPNWQEIVA AHEEERGVTVEQREEHQIPTISTSELRNESPRTY KGRHCKKFYTEEGCPYGENCTFLHDEQSRARES VAISVNPTVGGGTAYGANGPTQKPSNWKTRICN KWETTGSCPFGSKCHFAHGLAELHKFSGLDGEG RDSPMPAETKPSGMTTKTDTLTSPNLLTLPRQEY HHVGIPSQRSSSVPQAVGQRAAHKWKGPDKISR IYGDWIDDIE
64	sp_112710_rsfe	0.944036	0.93887	0.936349	unknown protein; INVOLVED IN: vegetative to reproductive phase transition of meristem; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G40640.1); Has 103 Blast hits to 103 proteins in 14 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 103; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MRMGTAFVVAAVSLFDEYANDLCYLEPGSCLPHP RYRKNVPPDSIQRRSSHDAENSLGKDGKEGLSD SKLVSQRSRTLKAAIQHYKPFKVLDWLFKSCEVN GMEFLNDGLINVKDIEECVVKGDCKKLGTALPAW CILKCLLASAKSDAPGLVVSDEADLTITNWPGDK MFQWLIEPLLVMKEQIRKLNLEDNEEACLRQAIM QCKNETPEDWDDIGFPSQDNVRRAQLQSVIRRL QGIVGSMTRMPTFRRRFQNLVKVLYVEAAKTGA LVSYIVENSTGSNTSQGDKLDQTSQETDEGKSD VANIV
65	sp_112720_jcmg	0.9467237	0.943704	0.938363	TINY2 (TINY2); FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: shoot apex, embryo, hypocotyl, leaf; EXPRESSED DURING: LP.10 ten leaves visible, D bilateral stage; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT5G25810.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MADGHERTEPETSQNDTTSTSSSQTPSGLPNNK RKRENSKDPVYRGVRMRAWGKWVSEIREPRKK SRIWLGTFNTPEMAARAHDVAALCIKGNSAILNF PELAKSLPRPASNSPRDVQAAAALAAAMVSPSP SQSPAVSSSSSSEVNGPEEELSQIVELPSLEDEGA DSAESMSELVLVDSVEEWAYSSWWTDEMMLRD FLYESTADISLADYGVGYTYY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
66	sp_113700_suxh	0.3068134	0.6057737	0.3383748	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: response to salt stress, N-terminal protein myristoylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: guard cell; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT4G27570.1); Has 5043 Blast hits to 5008 proteins in 241 species: Archae - 0; Bacteria - 56; Metazoa - 380; Fungi - 15; Plants - 4575; Viruses - 3; Other Eukaryotes - 14 (source: NCBI BLink).	MGGEKELRIVMFPWLAFGHFIPYLHLSNKLAEKG HKITLLLPNKARLQLESLNLHPSLITFHSITVPPLET LPYGTETTADISLDQHGELSISMDRTRPEVESFLS THKPDLVLYDMAHWVPEIAAKVGIKSVSYNVVCA IAVSHVRPSLPLPKGTAAHVPLPLSSVPKWSLNQ HGSSTPYFGEGITLLERSVISLSSADAIAIRTCREI EGVYCDRVAATFNKPVLVTSHALPDLELELSPLE TRWAEWLARFEPGSVIFCCLGSQHVLDAPQLQE LALGLEMTGLPFLMAVKPPVGCTSLEEVLPEGFN DRVSGRGVVHGGWVQQQQIMAHPSLGCFVTLC GSSSMWEGLVSESQLVLLPQLADQTLYAKLMAD ELKVGVKVEREENGWMTKRSLCEAIKSVMDEDS DISHVVRKNHAKYRSMLISPGFISGYIDNFIKDLQA LVP
67	sp_114580_drwu	0.3021846	0.5873217	0.3327153	Heavy metal transport/detoxification superfamily protein ; FUNCTIONS IN: metal ion binding; INVOLVED IN: metal ion transport; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Heavy metal transport/detoxification protein (InterPro:IPR006121); BEST Arabidopsis thaliana protein match is: Heavy metal transport/detoxification superfamily protein (TAIR:AT3G05920.1); Has 700 Blast hits to 667 proteins in 32 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 700; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MDDNKGILTVVGEVEPVPIFKALKKIGQTPEIVSV GPPKPQRRGPFYNCSSCRSYSDPQQKGWDTA MYPCSLPCNRSCCKECS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
68	sp_115130_uqnn	0.937821	0.947707	0.932838	starch branching enzyme 2.2 (SBE2.2); FUNCTIONS IN: 1,4-alpha-glucan branching enzyme activity; INVOLVED IN: amylopectin biosynthetic process, starch metabolic process; LOCATED IN: chloroplast, chloroplast stroma; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 13, N-terminal (InterPro:IPR004193), Immunoglobulin E-set (InterPro:IPR014756), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Alpha-amylase, C-terminal all beta (InterPro:IPR006048), Immunoglobulin-like fold (InterPro:IPR013783), Glycosyl hydrolase, family 13, all-beta (InterPro:IPR013780), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781), Glycosyl hydrolase, family 13, catalytic domain (InterPro:IPR006047); BEST Arabidopsis thaliana protein match is: starch branching enzyme 2.1 (TAIR:AT2G36390.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVYTISGIRLPSTVPSIHNKTSCFHGDIRRNSSFS LFLNNKSSASLSRKIYAKKSSYDSDSPSATLADS QKILVPSGEAEGSLSDESLTSDPNLNVPEGFHDF DVQTMKEGKESEADKHGSVAATRSEDKDVQNF VTRRTVPPPGNGQKIYEIDPLLKDHRTHLDYRYS MYRKLREEINKYEGGLEEFSRGYEKMGFNRSAE GITYREWAPGAKAASLVGDFNNWNSNADVMTR NEFGVWEIFLPNNADGSPAIPHGSRVKIHMETPS GIKDSIPAWIKFSVQAPGEIPYNGIYYDPPEEEKYI FKHPQPKRPKSLRIYESHIGMSSTEPVINTYANFR DDVLPRIKRLGYNAVQIMAIQEHSYYASFGYHVT NFFAPSSRFGTPEDLKSLIDRAHELGIVVLMDIVH SHASNNVLDGLNMFDGTDSHYFHSGARGYHWM WDSRLFNYGSWEVLRYLLSNARWWLDEYKFDG FRFDGVTSMMYTHHGLQVSFTGNYEEYFGFATD VDAVNYLMLVNDMIHGLFPEAIAIGEDVSGMPTF CLPTQDGGVGFDYRLHMAIADKWIELLQLRDED WRMGDIIHTLTNRRWDENCISYAESHDQALVGD KTIAFWLMDKDMYDFMAVDRPSTPLIDRGIALHK MIRLITMGLGGEGYLNFMGNEFGHPEWIDFPRG DQHLPTGKVVPGNNNSYDKCRRRFDLGDANYLR YRGLQEFDQAMQHLEENYGFMTSDHQYISRKSE DDRIIVFERGNLVFVFNFHWSNSYFDYQVGCLKP GKYKIVLDSDDPLFGGFNRIDHSAEHFTFEGWYD NRPRSFLVYAPCRAAVVYALVEDEAEAAHALVEE DSEPGHALVEDDPEPAMQALPVKDETKPAEE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
69	sp_115320_jnco	0.949668	0.92597	0.960168	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, kinase activity, ATP binding; INVOLVED IN: transmembrane receptor protein tyrosine kinase signaling pathway, protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat protein kinase family protein (TAIR:AT3G47580.1); Has 213556 Blast hits to 133070 proteins in 4776 species: Archae - 194; Bacteria - 22980; Metazoa - 68117; Fungi - 10236; Plants - 86850; Viruses - 328; Other Eukaryotes - 24851 (source: NCBI BLink).	MKPQPLLYTQAAVILLVVVIAMQPHPLRVANAISN ETDYMALLAIKSQLIDLHPNGVLSSWNHSLHHCL WEGVICGRKHKRVTGLKLFGRGLSGTISPFIGNL SFLNIIWLDSNSLHGSIPPEIGRLLRLQSLYLPNNT LSGGIPASLSSCINLQNLSLAINNLEGELPMELGA LSNLKWLWLHKNKLSGPIFKLILNLTSLVRIKGHT NAFTGPIPDNIPKLTNLTSLELSTNQLSGIIPPSISN LSSLQILDLSMNQLSGIIPPSIFNLSSLQILYLAYNQ LRGSIPPHIALTLPQLQYLQLAYNYFSGPLPISIHN LTSLRSIELSNNYFKGTISNLHFGHLYNLQRLVISN NNLSGDINFITKLVNCSKLESLDVGNNQFTGLVPK VLANLSTNLIEIWIEGNQVTGGIPVGLGNLNNLRIL SMKSSGLTGTVPRDLGELQNLEILDLSFNRLTGEI PSSFGNLSHLSNLNFYENKLQGRVPSSLGNCAN LLYLYLSYNQFNGSLPIELFAKTTNFIQLWLDQNH FQGPLTMEISKQINLEVFSMSNNKFFGSIPDVFSS LPALQELYMEDNFFHGLIPPSFASLKSLLKLDLSQ NNLSGPVPEYFATFPLIVLNLSHNNFEGRVPTKG VFANSSATSFVGNKNLCGGSSKLHLPRCVENER KKRRMSRALKLTIITASAFVGVLVMAICIWLYLKG QSKKRKSTSSSDALVKERFLKVSYDMLLKATDGF SQENLLGSGTFGSVFKGILDGKTVAVKVLNLQQR GGSKSFMAECEALRYIRHRNLVGILTACSSIDFKR NDFKALVYEFMPNGSLDKWLYKSGYLSLLQRVNI AIDVAHALNYLQCECETPIVHCDLKPNNILLDNDM VAHVGDFGLAKVLGQPLHPNQSSSIGVRGTVGY AAPEYGLGGEASPEADLYSYGILLLELMTKRRPT DNMFKEDFNLHMFAKAALPHQVLHIVDSTLMEDE SDEPDKRGPKPHEMLRKTEECMVSVIKIGVACSS HLPRDRMKISEAISELQKARNILLSPKHKRNLPRG ESD
70	sp_118860_ouyc	0.994609	0.959687	0.966049	uncharacterized protein	MHTMKPTVSNKAQPEGSIANARIMEECLAFISRY LNGIETKFNRMSRNDMDMQESLLFKLSVFQKKG NPLGKRTFKQLSFLDWKQAQLYVLMNCQEVQPF IGEYTTIHGPKPSLVDWFQSQIWKLYQERDPRVT AELLSLSRGPSKSVKSYQGYYVNGYRFHTKKRQ RRRKTQNSGVVVKGDEESGEKDFFGVLKEVIEL EYDAPNNGDKEPIVVLFRCVWFDVYREGRGIKR DKFGGISLNFKKFLGTNEPFAMASQVGQVYYVRF DDEAEIKYNEFKKLHEKEVGETGADNVSVDVAYE KVLGYRSGYARGLGKGHPVLSKDEKKGRAELEV TVVQLQNENNTMRAEFEHHKAETLRKENEAKKK QEELEQKLKLIMEKIGLESLT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
71	sp_125800_kzws	0.982317	0.967052	0.9902818	HXXXD-type acyl-transferase family protein; FUNCTIONS IN: transferase activity, transferring acyl groups other than amino-acyl groups, transferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Transferase (InterPro:IPR003480); BEST Arabidopsis thaliana protein match is: HXXXD-type acyl-transferase family protein (TAIR:AT3G30280.1); Has 2415 Blast hits to 2403 proteins in 134 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 58; Plants - 2353; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	MGEVNHEEVEIEIISIETIKPSSLLPPKTPPKTITLS HLDQAAPLYYYPLLLYYTNTTTTTPTSQIRVDITST LKTSLSKTLDKFHPIAGRCVDDSTICCNHQGIPFIE TKVDSNILDVMNSPEKMKLLIKFLPHAEFQDVTRP VSDLNHLAFQVNVFRCGGVIIGSYVLHKLLDGISL GTFFKNWSTIANDERVKDDDLVQPDFEEATIKAFP PRTATPMLPRNQQLPKAAEKPNNNPVKVLVTKS FVFDIVSLKKKMMFMAKSELVPKPTKFETVTGFIW EQTLSTLRNSGVEVEHTSLIIPVNIRPRMSPPLPR GSMGNLLKNAKAQANTSSSNGLQDLVKEIHSSLS QTTQKINTPPPPPPPPPPTTTATTIHSSLSQTTQKI NTPPPTTTTIHSSLSQTTQKINTTTTTTAEVILTKRK VDNPVTQNREGNYLFTSWCKIGLDEADFGFGKP VWVIPNDGRPPKVRNMIFLTDYRHPETGVEGIAA WITLEEKQMQCLKSNPEFLAFATPN
72	sp_126640_aupr	0.932137	0.9641969	0.9711087	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT3G61400.1); Has 8339 Blast hits to 8299 proteins in 1003 species: Archae - 0; Bacteria - 1128; Metazoa - 122; Fungi - 867; Plants - 4923; Viruses - 0; Other Eukaryotes - 1299 (source: NCBI BLink).	MSLESFTNGFQSNDDRLKEVKAFDATKMGVKGL IDSGVKCIPKIFIRPLDELSEDSKTPCVKLQVPTINL EGIDRKDRYNEIVKEVLDASETWGFFQVVNHGIS VELLDEMLQRVRMFHEQNTEVKKEFYTRETLKQ VVYNSNYDLYTSRSANWKDTLVVNAYYGHLDPK ELPEICRDVMLDYTNHVMKLSDILLELLSVALGLG PDHLREMEFKKGWSLVNHYYPACPAPDLTLGAS KHSDSAFLTILLQDDQIGGLQVLHENQWVNVQPKP AAFVVNIGDILQMVSNNKLRSVHHRVIANLIGPRIS VAFFLRGVLSSEKLYGPIKELLSKESPPVYREFTL SDFYSYFFSRPLNESGFEYFKV
73	sp_129380_nkfi	0.9514146	0.9444456	0.9869478	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein; FUNCTIONS IN: rRNA methyltransferase activity; INVOLVED IN: rRNA processing; LOCATED IN: cytoplasm; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: rRNA small subunit methyltransferase G (InterPro:IPR003682); Has 6840 Blast hits to 6840 proteins in 2435 species: Archae - 0; Bacteria - 5096; Metazoa - 2; Fungi - 0; Plants - 38; Viruses - 0; Other Eukaryotes - 1704 (source: NCBI BLink).	MLCLRSKLILSPFSFRTFIEHLPYNKPKSLNFTLNT RKSLNTAASVKSSHFETLNPHQKHQISLYVDALIQ WNQKMNLTAVTERSEVMERHIEDSLAILPPINES YTSHCDSSCESIKLVDVGTGAGLPGLILAIACPG WKVTLLESMNKRCLFLEHAIDVTGVSNVQVVRD RAENFGQTLEFRESFDIAVARAVAELRVLAEYCIP LVRKGGLFVAAKGHDPQEEIRNAERAMNLLGAS LLHQHLVESHGPFGQRTAIVCFKDRLTSRKYPRS PGTPAKMPL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
74	sp_135830_rgsg	0.968107	0.961157	0.976894	Leucine-rich repeat transmembrane protein kinase; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Malectin/receptor-like protein kinase (InterPro:IPR021720), Protein kinase, catalytic domain (InterPro:IPR000719), Leucine-rich repeat (InterPro:IPR001611), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat transmembrane protein kinase (TAIR:AT1G53420.1); Has 185540 Blast hits to 134573 proteins in 4786 species: Archae - 122; Bacteria - 17971; Metazoa - 51218; Fungi - 10628; Plants - 83318; Viruses - 446; Other Eukaryotes - 21837 (source: NCBI BLink).	MESSFHTVLIVFWVLFSLFGLSSELCQRTAAKEV EALREIAKELGKKDWNFNEDPCSNHTSWFSRGD PARQTYVNQVTCNCVSSRWHLTNISLKGQDLSG SLPKSLVKLPYLKTIDLSRNLLSGMLPDEWATMK LEFVSLMANQLSGQIPKFMGNITTLLYLSIESNSF SGTIPAELGKLVNLKHLFLNDNNLRGHLPIELTKL ASLTELRVSSNNFSGQIPNYFHHWKQLEKLQIQG SGFDGPIPSSISLLTNLIELRISDLNGEGSEFPSLE NLTGLSRLMLKSCKLYGTIPSYIQAMSRLKLLDLS FNKLNGSIPNELQFLNKLQNVYLSSNMLSGPIPS WITNKEDSSHIDLSYNHFDESSAPSSCNKETLNL YRSFSRGENLESSVCLKNFKCKARYSLHINCGGA ETTVAKTTYQKDDERGGAAKFIPRRAEWGFSSS GEFWFSQLPSDYIAKGNSAIRLKDSNLYMRARLS ATSLTYYASCLGKGNYTVTLHFCEIVFANNSYSSL GRRIFDIYIQDKLVRKNFNIVKEANGTNKIVKKVFP HIPVSDSLEIRFYYAGKGSTGIPVRGTYGPLVSAI SVQSDFKPPLDKQKIIRILSGVAILILCTTILGGFIL WRKHRSRDIASREQVLAGLDLQTGLFTFRQIKAA TNNFSADSKIGEDNSQLKLNWLTRQKICIDVARG LTFLHEESTIKIVHRDIKATNILLDGDLNAKISDFGL ARLDEGESTHISTRVAGTM
75	sp_140780_kjmt	0.94493	0.925374	0.943363	Serinc-domain containing serine and sphingolipid biosynthesis protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: membrane; CONTAINS InterPro DOMAIN/s: TMS membrane protein/tumour differentially expressed protein (InterPro:IPR005016); BEST Arabidopsis thaliana protein match is: Serinc-domain containing serine and sphingolipid biosynthesis protein (TAIR:AT3G24460.1); Has 814 Blast hits to 786 proteins in 182 species: Archae - 0; Bacteria - 0; Metazoa - 443; Fungi - 138; Plants - 166; Viruses - 0; Other Eukaryotes - 67 (source: NCBI BLink).	MAGERTDLRVSEAVYLEELSRTPQKKIDVSVEKR KSLKVRYYYGIIFLTTNFVAWLVRDYIQRVIPENH CKLGLXVRDYIQRVIPENHFLRTCGVGGHDCIQTI GVLRISFGCFIFFFLMFLTTLNTNKLQEVRNAWHS GEFARIGAGVFLALQLISVIQFIAWWNNYWMPDV KRKQSCSLGLFMSTVFYVASICGVVALYILYVPRS SCTLNIFFITWTAVLLIVMMLITLHSKVNRGLLSSG IMAAYVVFLCWSAIRSEPAGDKCSPQKQVTGHH DWITVFSFFIGICAIVMATFSTGIDSESFQFRKDEV EEEDDIPYKYGFFHLVFSLGAMYFAMLFINWDLN SSTRTWSIDVGWASTWVKIINEWFAATIYMWKLI SPVVRQAKIVDEGAIQPDQSC
76	sp_146060_hxqp	0.959415	0.926268	0.937799	uncharacterized protein	MNPFGVPISNETLKAMSRYKDKSITQEDRAREAI RLIHAEDKNLSALKHALKLKSDYGNGVSTLCLLYN ATGDTLELVQDQKQDWLGYIYKEEPPSSFENGQ WLAFLHAHPTSQAYGSEAARVYRGKNVKGDVR DFLVAWSTPWGSTQNSAYTEVREKDHFPQYWD YIKVSKLENANKITKDESDEYCASTVSIGGITSPEF IAVLKHKFSPLP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
77	sp_146920_mcah	0.98681	0.950642	0.930549	Prolyl oligopeptidase family protein; FUNCTIONS IN: serine-type peptidase activity, serine-type endopeptidase activity; INVOLVED IN: proteolysis; LOCATED IN: chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase S9, prolyl oligopeptidase, catalytic domain (InterPro:IPR001375), Peptidase S9A, oligopeptidase, N-terminal beta-propeller (InterPro:IPR004106), Peptidase S9A, prolyl oligopeptidase (InterPro:IPR002470); BEST Arabidopsis thaliana protein match is: Prolyl oligopeptidase family protein (TAIR:AT1G20380.1); Has 8127 Blast hits to 8050 proteins in 1368 species: Archae - 105; Bacteria - 3923; Metazoa - 339; Fungi - 43; Plants - 244; Viruses - 0; Other Eukaryotes - 3473 (source: NCBI BLink).	MLSRGPIVGYSNFQLCAGFLLFPRTRTTSLSSSL SPPNYKYSSNYFCCHLHKHKPTLLNSSHFLSPQT FSSMASASHKPLHYPLSRRDDSVVDDYFGVQITD PYRWLEDPDSEETKEFVENQVKLTDSVLGECET RDKLKEKMTELFDYPRFYPPFKRCDKYFYFHNA GLQPQSVLYVQDSLDGKPEVLLDPNQLSEDGTV ALNTYSVSEDAKYFAYGLSASGSDWVTIKVMRIE DKGVEPDSLSWVKFSGITWTHDNKGFFYSRYPP PNEGQISDAGTETNSNLNQELCYHVLGTDQSEDI LCWRDPENPKHSFGAEVTDDGKYLLLSTSESCE VVNKVYYLDLTTLPNGIEGFRGGKDLLPFTKLIDT FDAAYDVVANDDTVFIFRTNKDAPRYKLVRVDLK SPGIWTDVIPQSEKDVLESAYAVNGNHLLVCYLS DVKHVLEVRDLQTGSLLHHLPLEIGSVDDISARRK DSVVFFRFTSFLSPGIIYHCDLKSDVPELKIFREIV VPKFDRDEFQVDQVFVSSKDGTKIPMFIVTKKGIV LDGSHPCLLYGYGGFNISITPSFGVSRIVLLKHLG AVFCIANIRGGGEYGEEWHKAGSLSKKQNCFDD FISAGEFLVSYGYTQPKKLCIEGGSNGGLLVGACI NQRPDIFGCALAHVGVMDMLRFHKFTIGHAWTS DYGCSDKEEEFHWLIKYSPLHNVRRPWDLSSDR TAQYPSTMLLTADHDDRVVPLHSLKLLATLQHVL CTSLENSPQKNPIVGRIECKAGHGAGRPTQKMID EAADRYGFMAKMVNASWTD
78	sp_147470_tnri	0.990207	0.935739	0.944002	uncharacterized protein	MANPFGSVVDDEKLDEMERYIGKTKTQEDRARE AMHLINEDDKNSKAEAYAESVKEYYGSGVSTMC MVYNATGDTLTYVTDNDWYGFISRTPYPTEIGNG QWAAFQHVHNTGASSGSEAAVVYRGKNKDGHE RDFMLSWSTPWGPWYKNKAYCEMGGVDSFQS RWDDIYDKLNNSGYSDHVDRDGVKIDVDTATGG APIFHATIKIPFSS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
79	sp_148210_yuyx	0.9883242	0.9581131	0.9403448	PDI-like 1-1 (PDIL1-1); FUNCTIONS IN: protein disulfide isomerase activity; INVOLVED IN: in 6 processes; LOCATED IN: in 8 components; EXPRESSED IN: 29 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Disulphide isomerase (InterPro:IPR005788), Thioredoxin fold (InterPro:IPR012335), Thioredoxin domain (InterPro:IPR013766), Thioredoxin, conserved site (InterPro:IPR017937), Thioredoxin-like subdomain (InterPro:IPR006662), Protein disulphide isomerase (InterPro:IPR005792), Thioredoxin-like (InterPro:IPR017936), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: PDI-like 1-2 (TAIR:AT1G77510.1); Has 31573 Blast hits to 17834 proteins in 2924 species: Archae - 337; Bacteria - 15428; Metazoa - 5451; Fungi - 1644; Plants - 2422; Viruses - 29; Other Eukaryotes - 6262 (source: NCBI BLink).	MALKIGGLLEVCLVVLLATIGGIQVEAEKEEYVIT LDHSNFTDVVSKHDFILVEFYAPWCGYCKQLAPE YEKAAATLSKHDPPITLAKVDLNEDTNKEHFKSQ FEIRAFPTLKFLINGGKIVHDYKNPPQSADDIVQQ VKKLLGPVSAEIKSVNDASKIIDQKKVFIAGVFPKF EGEEYGNFIKLAEKLRSDYEFGHTLDAKVLPRGD SSVKGPVVRLLKPFDDLFADTKDFDLAALEKFVQ EEDVPSVIPFTMDPSNTYLRKAFSGENTKIMMVL DFKSENAAALKSRFSDAAILLKGDNKYFLVGDLE SSKDLLKHFKLEDSHAPLVFIHKTEAQKYLKTNVE PDQLAPWLKKFLDGKLEPFLRSQPIPEENNEPVK VVVLDSLQDMVFNSEKEVFLEFYAPWCGHCKSL APILDEVAISYEKDANIMIAKFDASANDAPKDTFNV KGFPTMFFITKSKEIVEYNGGRAKEDIVGFIEKHR TTEKNKSEAGIKSEPEKENNHVDETVASKDEL
80	sp_148230_dgra	0.9984452	0.9878422	0.9753379	cytochrome P450, family 72, subfamily A, polypeptide 15 (CYP72A15); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: response to cycloheptenone; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 72, subfamily A, polypeptide 13 (TAIR:AT3G14660.1); Has 31875 Blast hits to 31741 proteins in 1631 species: Archae - 65; Bacteria - 4943; Metazoa - 11298; Fungi - 6149; Plants - 8001; Viruses - 3; Other Eukaryotes - 1416 (source: NCBI BLink).	MKESSTLRAQALNKPIPFTHDYYQRIQPFTHQILN NSCAGAGKNIYTWLGPVPTILITQPEIIKDAFNRM NNFQKPRLNPYTQILSTGLPNYEGQKWAKHRKL LNPAFQLDKLKLMIPAFETCVTDTLNKWEKLVSK TGSSEVDVWPHFTTLTGDGIARAAFGSSFEDGR RIFDLLKEQKDLVISLLKYSYIPGFKYLPTKGNKM MKETENEIKPLLTNLIHKFRKKAMEAGEAPKDDLL GMLLESNANEARQVNENESGSRKRQSDITMSFH EMIDACKLFFLAGQETTSVALTWAMILLAKHQDY QTRAREEVLATFGTKTPDFDGVHNRLKIVTTIVHE VLRLYPPIPATSRRAHEREIKLGDLVIPQGVGVSF SILHAHLNPEIWGDDAKEFKPERFSEGIAKATKG NNSYFPFGWGPRICIGQAFALIQVKMALSMILQR FSFELSPSYIHAPTSLLALQPQHGAHVILHRL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
81	sp_148240_fnwi	0.948312	0.963263	0.96873	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: N-terminal protein myristoylation, metabolic process; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT4G27560.1); Has 4908 Blast hits to 4874 proteins in 244 species: Archae - 0; Bacteria - 41; Metazoa - 215; Fungi - 17; Plants - 4624; Viruses - 3; Other Eukaryotes - 8 (source: NCBI BLink).	MCDDKKSSVLSIAFYPWFALGHLTSFLRLANKLA QNGHNVSYFIPTNTLPRLLPHNHYPGHLTFIPVTV PPVDGLPLRAETTNDVPSSAIHLLMTAMDLTRDTI EAHLVSIKPDVVFYDFAYWIPDLARKHGFKSVLYI TSYIARCAYFAPDLKSGHQSTGAEIIAPPPGFPSQ HFRMQAHEAETVADVGKEQDGLQGITISERMRIA FGKCDAIGVKSCKEMEKVYIDYCEKIFGKSVLLAG PMVPKTPSSKLDEYFDGWLTGFGAATVIYCAFG SECVLEINQFQQLLLLGLELTGRPFLVAMKPPKKY ETIESALPEGFEKRTKGRGIVHEGWVQQQLILQH PSVGCFITHCGVGSLSEAMVSKCQVVLMPQAVD QFINARMMSLELKIGVEVEKREDDGLFTKEAVHK AVSLVMEEESEVAKEMRVSHDKWREFLLQEGLE DSYISSFIQSLRQLTIG
					serine carboxypeptidase-like 18 (scpl18); FUNCTIONS IN: serine-type carboxypeptidase activity; INVOLVED IN: proteolysis; LOCATED IN: endomembrane system; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Peptidase S10, serine carboxypeptidase (InterPro:IPR001563); BEST Arabidopsis thaliana protein match is: serine carboxypeptidase-like 7 (TAIR:AT3G10450.1); Has 3653 Blast hits to 3603 proteins in 397 species: Archae - 0; Bacteria - 261; Metazoa - 616; Fungi - 829; Plants - 1537; Viruses - 0; Other Eukaryotes - 410 (source: NCBI BLink).	
82	sp_152910_apyy	0.938871	0.973287	0.955189		FPGNLPFYLETGYVRVGENDEVQLFYYFIKSERD PEKDPLMLWLTGGPGCSALSGLVLEIVSRCGLN STMSGSRKTDGHRWWFWSCVGGCWSLTAHRR RENEKEMERCPLNFNTSACNWHSKSPTFQLNPS SWTKVASIIFLDSPVGTGFSYATNPEGYYTDDITS SRHIYQFLRKWLMDHDGFKDHPLFISGDSYCGKI VPIVVQEISNGNRAGFEPAMNLQGYVLGNPYTRK EDDNKSKFYYAHRVSLLSDDLYESATRINCNGTY VDVDVDKDNVKCTRILQAISYNLDPLFDAHVLKPK CISHQTWCQVYFSCSSDMIA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
83	sp_157560_hiix	0.943223	0.922952	0.948664	purple acid phosphatase 26 (PAP26); FUNCTIONS IN: protein serine/threonine phosphatase activity, acid phosphatase activity; INVOLVED IN: phosphate ion homeostasis; LOCATED IN: vacuole, plant-type cell wall; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Purple acid phosphatase, N-terminal (InterPro:IPR015914), Metallophosphoesterase (InterPro:IPR004843), Purple acid phosphatase-like, N-terminal (InterPro:IPR008963); BEST Arabidopsis thaliana protein match is: purple acid phosphatase 12 (TAIR:AT2G27190.1); Has 1843 Blast hits to 1825 proteins in 387 species: Archae - 3; Bacteria - 583; Metazoa - 186; Fungi - 81; Plants - 765; Viruses - 0; Other Eukaryotes - 225 (source: NCBI BLink).	MGSADRAQRKEEIGLNGADNYGPWTVVTNKRS TKLKLGQGRSANSSYKGANSKINEYRGNKSTPVI NVHQFPTGSNARYMKGQNITQLERINSSANLDFQ KNNNDGDFKGYSASKKSKGLEEELGHVPQALTH LVGGSDSPVSSLPFQINSAPVNQSNTPTSHSSST LSPNLHNISVSVSSFNDQPNSISTSIGGKCQSIS TNADHISSSDNLPAQRSSELDSQRAKAKHAESS GGERRSIEPDTVQNPQSVPTTSNDEPMFIGDGE KSGEGNCMHAGEYSGISAEERDNSTQEANASKG GSYELTSHASFTSSYYRLDPPSTDIPLDSEFFAAP EAYNAPQQVHITQGDYDGKAVIISWVTPYLHGSH TVYYGTTKGKYTSKAATNTSYSFNVYKSGQIYH CLVDGLEYNTKYYYRIGSGPSNREFWFQTPPEV GPDSPVTFGLIGDLGQTMNSLSTLQHYMKGEGQ TVLFVGDLTYADRHIDEAPHEGGLRWDTWGRLV EPSTAYQPWIWTCGNHDIDYFPKLGLLGEFMQF RERYPSPYKASESTSPLWYAIRRGPAHVIVLASY SGFVKYTAQYRWLDNELKKVDRKKTPWLIVIMHV PLYNSNTAHYMEGESMRSVFEQWFVDYKVDVV FAGHVHSYERSHRVSNLKYNISNGACSPVLDESA PFVVTIGDGGNSEGIAGEYISPQPHYSAFREPSF GHAILDIKNKTHAYYYWHRNQDGNKVTADSFVLH NQYWTNYV	
84	sp_158080_oxzh	0.93227	0.88511	0.965032	monogalactosyldiacylglycerol synthase 2 (MGD2); CONTAINS InterPro DOMAIN/s: Monogalactosyldiacylglycerol synthase (InterPro:IPR009695), Glycosyl transferase, family 28, C-terminal (InterPro:IPR007235); BEST Arabidopsis thaliana protein match is: monogalactosyldiacylglycerol synthase type C (TAIR:AT2G11810.1); Has 1572 Blast hits to 1572 proteins in 585 species: Archae - 0; Bacteria - 1385; Metazoa - 0; Fungi - 0; Plants - 105; Viruses - 0; Other Eukaryotes - 82 (source: NCBI BLink).	MVMSVIAPKRSSSLSSVFSRVGGYYSKKRCSSQ FYPSSSSGISSSSSSSMVIDEMEDDEDGSTMEMV QLGADRTKNVLILMSDTGGGHRASAEAIRDAFNL EFGDEYNIIIKDVWKEYTGWPLNDMERQYKFMV KHVQLWSLAFHSTSPKWIHSVYLAAIAAFYAKEV EAGLMEYKPDIIISVHPLMQHIPLWVLRWQGLQK KVIFVTVITDLNSCHRTWFHPAVNRCYCPSEEVA KRASLDGLEDSQIQVFGLPVRPSFARAVLSKDDL RVELEMDPDLPAVLLMGGGEGMGPVQKTAKALA ESLFDKKLEKPIGQLIICGRNKSLAST LESKEWKI PVKVRGFEKQMEKWMGACDCIITKAGPGTIAEA MIRGLPIILNDYIPGQEKGNVPYVVENGAGVFTRS PKETARIVAEWFSTKGDELRKMSENALKLAKPEA VFEIVRDLHELADQRGPLANLPYMLTSSFTSLI	
85	sp_166530_hghx	0.870423	0.963423	0.930782	F-box family protein; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364); Has 68 Blast hits to 68 proteins in 20 species: Archae - 0; Bacteria - 0; Metazoa - 4; Fungi - 0; Plants - 59; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MLVCKQWKDIAEDDYLWKCLCARRWPSLCKRP SPSTMKYYKLYGTFHKRQRRRDLPPPRLSFNDL DFYIDIWAEGQSIFSEVVSGPVLHAGIKSRPIGISD IVRCHLDGPEYRMMLPVEPRFSVPLGQTASVSVL VERKDFNKFACIINKSLFDYVDRTSFLAIAFDYLDF APIHPFISGIRAWISLLFTDGENDRVADVFGIEMD FCDAASNETEVLWLLDMLDWK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
86	sp_168890_mkyf	0.956338	0.99422	0.975675	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G14850.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MVVRNMDNIKMPGDRHFDRIDILEIKLQISRKLGN LRAEKYFGLLDRFFGLKLCKSEFDKLCIVAVGRE NISLHNRLIKSILKNACLSKSPPQKRSKLETPIAEK VANGFQRSCLQSLCRDALPPSPRRGRIPGFRDR KSKDRQSPLGPNGKSNTVEACKDLGHKTLEQQS ATELLSLGSRPIEVNSVEDGEEVDQAAISPGIHSR SPVTPPLGIPWNGKGNRKVLSSLLSSSGNLESC HGNGELPNTSCLRKRLEQKLEKEGLGISQDCVNL LNDGLDTYLKRLIKPSLELANSRNGNKRQKVGGE DVHVHGLNGLWPVRYLQSPPVSSSVTILDFRTA MELNPEVLGEDWPTQLEKVCLREFESEYSD
87	sp_168980_dmck	0.991104	0.991699	0.990633	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Golgi apparatus membrane protein TVP15 (InterPro:IPR013714); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MRNAEGEGEISQPRNSPPDSPSPHQSRPTSNTK PDAKLRRRNAPFLIVCKCFSVITALSAILCIAVNILS AFRSFKNGSDVFDGIFRCYAVALAIFVVLAETEW SLIFRFWKVLEYWPGRSMLQIFVAVMTRAFSDFI GKQRDIVLLQSIASYLLLSCGVVYLVSGILCIGVIK RSLEEKEISRDQAVKDLEELEKKRHELEALLIEEE NV
88	sp_169540_akgo	0.953251	0.934449	0.936047	EARLY LIGHT-INDUCIBLE PROTEIN 2 (ELIP2); BEST Arabidopsis thaliana protein match is: Chlorophyll A-B binding family protein (TAIR:AT3G22840.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MAMSMQVQALLSNSVGAFSTGSSKAYPVGFNP RIAKRVNFGVKCMAEEKVGGEPKLDSGLPKPNP TVLSSPTSIPTPRPQPKKVSTNFTDLFAFSGPAPE RINGRLAMIGFVAAIAVELSNGTDLFSQISGSGSQ WFLGTSILFSAASLVPLLQGSRPEAKTEGLMNAN AELWNGRIAMLGLVALAFTEFVTGSALV
89	sp_170320_dmqi	0.989012	0.988954	0.996891	UDP-glucosyl transferase 73B5 (UGT73B5); FUNCTIONS IN: quercetin 3-O-glucosyltransferase activity, UDP-glycosyltransferase activity, UDP-glucosyltransferase activity, transferase activity, transferring glycosyl groups; INVOLVED IN: response to other organism; LOCATED IN: endomembrane system; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-glycosyltransferase 73B4 (TAIR:AT2G15490.3).	MELSNPTTTPTLNATQPLRGYFIPLITVPSHISNLV DIAKLFSSRGVHVTILTTHHTSLRFKQSIHDWGFK IDLHIVDFPFREVGLPEGVENYSDATPEQASQLF QAFMMLQKPMEDAIRAAKPDFIVSDRYYHWSTD LARELAIPRLIFHVRCYFALCAAEVVAKFAPHEKV ESDTDLFFLPDLPDTIHMTRLQLPEWIQTRNMFT VLNERMDEADRECYGVIVNSCYELERAYADFYR SNLGRRAWCIGPYPVHCDKVGKKKGDDSKKHS CFEWLDKMGEGEVIYVSFGTLSCFSPAQISELAT ALEMSGHPFIWVVRNGEKLLPDGFEERITEQDK GVLIKDWAPQVKILEHPAVGGFLTHCGWNSTVE SLAAGVPMVTWPLGAEQFFNEKLISGVLKVGVE VGSEKWSRGIVPNTDMIEKDKIERAIKELMSKEP EAEERRQKVKELSKAPRNAVEEGGSSRNNLSDLI EKLQRLKANEISVSTNSE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
90	sp_170920_oudh	0.969733	0.941828	0.985877	Uridine diphosphate glycosyltransferase 74E2 (UGT74E2); CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT1G05675.1); Has 8397 Blast hits to 8309 proteins in 513 species: Archae - 0; Bacteria - 605; Metazoa - 2499; Fungi - 31; Plants - 5061; Viruses - 123; Other Eukaryotes - 78 (source: NCBI BLink).	MLVYDAYFPWALDVAKDHNILAAAYFVSSCAYVA SFYPMFLEEFGSDDQHPVVAAESVPGLLVELPS REEMERYAPKCAQSSSSDDKPDTGKKPLHPVYQ MVVSSITTLHLADFVLFNSFDHLEHQVVKWMTNL WRVKTVGPLLPSAYLDKRIESDVNYCVKPYKPNN EACMSWLNAKQVASVVYVSFGSVAKLSVEQISEI AKALKQIPSSFLWIVREAEQEKLPNDFITETSEKG LVMSWCPQLDVLAHEAVGCFITHCGWNSVIEAT SFGVPMLGMPQFMDHFLDAHFLEKVWGVGIRVK ADEKNFVTCDEIKRGLEKIIYGERGNKIKENATKW KELAKEAVGEGGSSDNSTGEIIKWLASS
91	sp_170930_hjgq	0.963787	0.950346	0.977457	Uridine diphosphate glycosyltransferase 74E2 (UGT74E2); CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT1G05675.1); Has 8397 Blast hits to 8309 proteins in 513 species: Archae - 0; Bacteria - 605; Metazoa - 2499; Fungi - 31; Plants - 5061; Viruses - 123; Other Eukaryotes - 78 (source: NCBI BLink).	MTGKGRTMEVIMMPFHHQGHLTPMLQFAKRFA WKGAGSIRITLATTLSTAQNMTNSKNNNNNNDY DFLTVESIYDDTDDSQLKFMGRMGKFKSEASLQL GRLITTKSIDNNKCMLVYDAYLPWALDVGKDHNI QAAAFFVQACAYMASFYPMFLEEFGSDDQHPVV AAAKAESVPSLSVELPSREEMERYAPKCAQSPS SDDKPNTVKKSLHPVYRMVVSSITTLHLADFVLIN SFDHLEHQLDVLAHEAVGCFITHCGWNSIIEATNF GVPMLGMPQFMDQFLDAHFMEKVWGVGIRAKA DEKNFVTCDEIKCGVNEIMYGDKANMIKENAAKW KDLAKEAVGEGGSSDKNIDEIINWLASS
92	sp_171760_xdak	0.959305	0.959574	0.981244	cysteine-rich RLK (RECEPTOR-like protein kinase) 25 (CRK25); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Protein of unknown function DUF26 (InterPro:IPR002902), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: cysteine-rich RLK (RECEPTOR-like protein kinase) 10 (TAIR:AT4G23180.1); Has 124291 Blast hits to 122724 proteins in 4390 species: Archae - 112; Bacteria - 13863; Metazoa - 45275; Fungi - 10764; Plants - 35367; Viruses - 449; Other Eukaryotes - 18461 (source: NCBI BLink).	MLVLLILVVFVINASNAVPLTLDYIIWDCYNSTTYP PNSSYETNLNHLLSTLAPQTSNSKHGFFTTKLNT TQNQDHPVYGLASCRGDINPNDCRDCVRQAST QVIRACPEQTQAVFWMEQCMLRYSNRSFSHLLQ ESPNKVRFNTHNVTNNVNDWIKLLNDTLIEVTSK AASDKSGNKFATKRVWFSPSQKYLYTLAQCNPD LTVADCAECFSLARSQYDGMAQRGGLVLTPSCN IRYQVKQFYGENSNYSITPLPTLSLPPPPNHHISTI TYTGKKKQTTNVVVAIVVPIVILGVLISFSTYFYTK KAKKRRSLDLKNVLDDLTTAESLMFELEILEAATN NFAFIRKLGQGGFGSVYK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
93	sp_172300_mjoa	0.992201	0.967157	0.944833	RNA-dependent RNA polymerase family protein; CONTAINS InterPro DOMAIN/s: RNA-dependent RNA polymerase, eukaryotic-type (InterPro:IPR007855); BEST Arabidopsis thaliana protein match is: RNA-dependent RNA polymerase family protein (TAIR:AT2G19910.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MASVPPELPPLPEAVEEQLGLICHEKSLLPADAV ARRYLSEVGEQESFRILNIIRNSGEVRKLSGFIVHL VKNSRSPRSPLPASPTISPQGNTTSGKKRKMES DSPGGLSSVRRVCIFNSPTKIVERVDPPKSELDQ STVEEVCILNSPTKIEEKLHLLKSETDRVLDALEKL EFRKTFLILNYKGCAKLEDAVYGLSADKILSYQSL GMRDFENEIWNDFGRNICQQKERINKGPYLNKA ATLLQKTVEDENVLMVKFMEQETKSLDTESMMMQ SPAFNKVSEEGIFVGRKCYRFFVFKDGGKEEKKK NPTSSPVKCYFVCTEGFASLLKQPPHLLKNMSV QKARGLFMHIHKATNLSNYMARFSLILSKTITLEL NYDVLRVEVVEDVPCRDGNGDIVYDKDDKAMIFT DGTGFISEDLAVRCPHNCFNGYVNNDENVQAML RPDGVKIIFYCGSLGSLEAFRLPHNLSGLTELKSL PGVPPLLIQFRMFYKGQAIKGTVLVNRKLPPNTIQ VRRSMLKIKHDEKLSGAPTVNSFEVVATSNKPKD AAVSKNLIALLHYGGVPKDVFLSMLWEAITKPLSV LYKTKAAVRGNVADVYSITESEFCSLPFCHWQM KVKVLFFGNSRLCLSVGLSLSLRTSVIPLFGLTT
94	sp_174820_kxmh	0.946332	0.944918	0.944835	YELLOW STRIPE like 3 (YSL3); CONTAINS InterPro DOMAIN/s: Oligopeptide transporter OPT superfamily (InterPro:IPR004813); BEST Arabidopsis thaliana protein match is: YELLOW STRIPE like 2 (TAIR:AT5G24380.1).	MRSDVELREIGREDFEEEDKVQDEIKRVPPWQK QITIRGIIASILIGVIYSVIVMKLCLTTGLVPNLNVSA ALLAFVFLRTWTSFLQKLGIVTVPFTRQENTIVQT CAVACYSIAVGGGFGSYLLGLNRKTYEQAGVDT QGNAPISFKEPGVGWMTGFLFVSSFVGLLALVPL RKIMIIDYKLRYPSGTATANLINGFHTSRGENEPR VEVHGFVKCLLISFLWSFFQWSYSGGNQYGFPQ FPKFGLKAWENSFYFDFNMTYIGAGMISSHLVNL SLLLGAVLSWGVMWPLIGDMKGKWFPSSLPESS MKSLNGYKVFISIALILGDGLYNFFKTLFFIIRSIHD RLKDKKLSSCPDSKKQALDDLVRNEVFMREGIPI WLACFGYTFFSIISIVVIPKVFPELKWYYVVVAYM LAPSLSFCNAYGTGLTDINMAYNYGKVALFVLSA VAGKDSGAVVGLVGCGLIKSIVSISSDLMHDFKT GHLTLTSPRSMLLSQAIGTAVGCVVAPLTFFLFYN AFDVGNPDGEYKAPYALLIYRNMAILGVQGFSALP QHCLQLCYAFFGFAVVANLVRDTSPKKIGKWVPL PMAMAVPFLVGGYFAIDMCVGSLVVFAWHRLNS QKASLMVPVVASGLICGDGLWILPSSILALVKRNP SIR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
95	sp_175360_idec	0.952231	0.953457	0.963454	Glycosyl hydrolases family 31 protein; FUNCTIONS IN: hydrolase activity, hydrolyzing O-glycosyl compounds; INVOLVED IN: carbohydrate metabolic process; LOCATED IN: apoplast, vacuole, plant-type cell wall; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 31 (InterPro:IPR000322), Glycoside hydrolase, catalytic core (InterPro:IPR017853); BEST Arabidopsis thaliana protein match is: alpha-xylosidase 1 (TAIR:AT1G68560.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKKKIPSLALGILLVFLLQYLVAGISTSENDPEGVI GYGYKVKSVKVDSGTRRSLTALLQLVKNSSVYG PDIQLLSITASLESNDRLRVRITDAKHRRWEIPDNI LHRHQPPPPPPHSLSSLYRTLLSSPTTNRRKILLS HPNSDLTFSLINTTPFGFTISRKSTHDVLFDAITPD PTNPNTFLIFKDQYLHLTSSLPGTRAHIYGLGEHS KPTFQLAHNQTLTMWAADIPSSNPDVNLYGSHP FYMDVRSSPVAGSTHGVLLFNSNGMDVEYTGNR ITYKVIGGIIDLYFFAGPSPAQVVEQFTRVIGRPAP MPYWAFGFQQCRYGYHDVYELQSVVAGYAKAKI PLEVMWTDIDYMDAYKDFTLDPVNFPLDKMKKF VNNLHKNGQKYVVILDPGISTNKTYETYIRGMKH DVFLKRNGKPYLGSVWPGPVYFPDFLKPSALTF WTDEIKRFLNLLPVDGLWIDMNEISNFISSPPIPG STLDNPPYKINNSGVMLPIINKTIPPTAMHYGDIPE YNVHNLFGYLEARVTRAALIKLTKKRPFVLSRSTF SGSGKYTAHWTGDNAATWNDLVYSIPSMLDFGL FGIPMVGADICGFLGNTTEELCRRWIQLGAFYPF SRDHSSLGTTYQELYRWESVAASARKVLGLRYT LLPYFYTLMYEAQLNGIPIARPLFFSFPDDIKTYGI SSQFLLGKGVMVSPVLKPGVVSVTAYFPRGNWF DLFDYTRSVTASTGRYVTLSAPPDHINVHIQEGNI LAMQGKAMTTQAARKTPFHLLVVMSDCGASFGE LFLDDGVEVTMGVNRGKWTFVKFIAASAKQTCII TSDVVSGEFAVSQKWVIDKVTILGLRKGTKINGY TVQTGAVTRKGDKSKLKSTPDRKGEFIVAEISGL NLLLGREFKLVLH
96	sp_177730_zafo	0.98274	0.983322	0.993142	Putative membrane lipoprotein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: male gametophyte, flower, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis; BEST Arabidopsis thaliana protein match is: Putative membrane lipoprotein (TAIR:AT5G17590.1); Has 36 Blast hits to 36 proteins in 9 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 36; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAVPEDSEAGGWYIGIISFLLFIFTVSGAIFLVLYM IQPTSITHSWYPVIGGALVCFPWFFWFLTLLYRIV SRACGFRMSCFGTSGVDHPDEVIPSGVSAVNNA SIGTNNGVVDGGNPVIGDLENLNGNENGNENEN GNRASGVRLQPCSPRRVHFGGVVVLGEDDDGG GGAVGNQESPKSLSSSSSSSSSSSSSSCDESEMPL KLCMAS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
97	sp_178160_fpjg	0.997436	0.98747	0.993035	Zinc finger C-x8-C-x5-C-x3-H type family protein; FUNCTIONS IN: zinc ion binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, CCCH-type (InterPro:IPR000571); Has 904 Blast hits to 850 proteins in 161 species: Archae - 9; Bacteria - 40; Metazoa - 431; Fungi - 54; Plants - 83; Viruses - 2; Other Eukaryotes - 285 (source: NCBI BLink).	MMVETKPFKTKLCVLYGRGRCARETCTFAHGDS ELRRFSASLPDRRIDRDRDNDLRSKLDRKRSPLRRC SPERDARGSHAYKRISPSHPSKRRKLRTHEYLEG PNGFSEGSKNEEGSADSRKDKMHVSGESRKML EDELKQIQYEVDVLFNQKSQMEIELEERVQEADS LSSKIQDLEAQLAGEKDKSKRVDSKIKKLVRAHV HHTRLQDEVKRSQLQLEQLVGQLVRDATPFSAIE KLNVNINHEEKVLELKDSVPSVRKQPHSSMEASQ GLKSARGNGVRVLGSSRNPLVRIDCQQEAEIRNK RKDVAGPYTESNPRMRNNASKHFDPLDKVKSLE SSLVAPPTSMAAHAVDDFIELMEAEENPEVLGTP TNFGSSLPFLLPPLPTMSQNNYFQDEGNDANVD VDGHEEDIEEVDIV
98	sp_181290_kwoe	0.98247	0.95837	0.978338	low-molecular-weight cysteine-rich 69 (LCR69); FUNCTIONS IN: peptidase inhibitor activity; INVOLVED IN: defense response; LOCATED IN: plasma membrane, plant-type cell wall; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Gamma thionin (InterPro:IPR008176), Knottin (InterPro:IPR003614), Gamma Purothionin (InterPro:IPR008177); BEST Arabidopsis thaliana protein match is: low-molecular-weight cysteine-rich 68 (TAIR:AT2G02130.1); Has 400 Blast hits to 400 proteins in 82 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 400; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MERSSRVFSVVLLMLVLVLSTEIGTKVAEARICES ASYRFKGICVSRSNCANVCKNEGFPGGRCRGFR RRCLCYKHCG
99	sp_183370_xrwc	0.963244	0.96138	0.977536	60S acidic ribosomal protein family; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translational elongation, response to cold; LOCATED IN: cytosol, cytosolic ribosome, ribosome, nucleus; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal protein 60S (InterPro:IPR001813); BEST Arabidopsis thaliana protein match is: 60S acidic ribosomal protein family (TAIR:AT2G27710.4); Has 1851 Blast hits to 1850 proteins in 328 species: Archae - 0; Bacteria - 0; Metazoa - 772; Fungi - 418; Plants - 417; Viruses - 0; Other Eukaryotes - 244 (source: NCBI BLink).	MKIVAACLLSSLGGKACPTLDDIKKNLRSGPFSR HFFHNGSLIDEDII

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
100	sp_185520_rkzu	0.936016	0.990085	0.955541	ZIP metal ion transporter family; FUNCTIONS IN: metal ion transmembrane transporter activity; INVOLVED IN: metal ion transport, transmembrane transport; LOCATED IN: endomembrane system, membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc/iron permease (InterPro:IPR003689); BEST Arabidopsis thaliana protein match is: ZIP metal ion transporter family (TAIR:AT3G08650.2); Has 3662 Blast hits to 3630 proteins in 1201 species: Archae - 124; Bacteria - 2163; Metazoa - 678; Fungi - 116; Plants - 147; Viruses - 0; Other Eukaryotes - 434 (source: NCBI BLink).	MDSQVLVALALSLVGGLSTSLGALFVVLSDAPNL KMLGLLQGFAAGLMLSISFFDLAHNAINSIGFLKG NIWFFSGVIFFAIIANFIPEPSLAPTSNVKSKKRAN GEDKGKDLMKKHRRQVMYSGIITAIGISLHNFPE GMAVFLGSMKGLRVGLNLALAIALHNIPEGVAVA LPVYFATQSKWQAFKIASLSGLAEPLGVIIVAYLF PTSLNPEILDGLLGAVGGVMAFLTLHEMLPLAFD YAGQKQAVKAVFFGMAFMSASLYFLEISLPADLS L
101	sp_187550_ghsy	0.966194	0.945592	0.941083	ATP binding;nucleic acid binding;helicases; FUNCTIONS IN: helicase activity, nucleic acid binding, ATP binding; INVOLVED IN: biological_process unknown; EXPRESSED IN: embryo; EXPRESSED DURING: D bilateral stage; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885), DNA/RNA helicase, C-terminal (InterPro:IPR001650); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT1G12300.1); Has 84744 Blast hits to 32036 proteins in 2584 species: Archae - 123; Bacteria - 11324; Metazoa - 2691; Fungi - 2451; Plants - 63733; Viruses - 4; Other Eukaryotes - 4418 (source: NCBI BLink).	MANKLGVWQKASAFINGGNNTHRLSTSLSINYLL SASSYFHSHTNFHPFVENEYEVDSELFLESVRQQ CQMGFRKLDDAVSLCNRMICMNPLPSIMLFTLILS SMTKIKPRVFLNYSTVIHLSNHLEWLGVSRNLYSF NILVNCYCHLGHVNFGFSILGRIIKVGYCPDIVTFS TLINGLVQNGLLEQATELLDKIVKLGFQPNIVTYG ALIKGLCRIGNYANALNLLVTMRSGFFLCKPDAV MYSTIIDSLVKDKLLTQAFDLFTEMKSKGDQPDVF TYCTLVRGMCIAGKWDDAKCMLNEMLESNIDPD VETYNMLVDMCCKSRKIREAEAILGLMTKRGDAP STVTYSSLLDGYCSCGRLDDASEVMDLMVKYGC MPDVVAFNTLLKGCCKSKKLDKALDLVQNMHLR GLAPNVVTYNTLMDALCKDNRIHCASQLLKEMEA AGLKPDAVTYSSLLDQLCKSSQIDEALEMVRQM GIKGLVPDIVIYSILIDGLCGVGRYQEARDIFFSLLG KGLKPNQYTYNALIKGLCKKGLMNDAIDLLKTME DNGCSPNDCNYDTIIRGLLSANDNEKALEFVFIKK SKGFTLNAHTNSLFNASLSDHDVSKCQQSFASE VSVEPEA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
102	sp_190300_zfwy	0.98557	0.977328	0.9808502	squalene synthase 1 (SQS1); FUNCTIONS IN: farnesyl-diphosphate farnesyltransferase activity; INVOLVED IN: sterol biosynthetic process; LOCATED IN: endoplasmic reticulum, plasma membrane, endoplasmic reticulum membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Squalene/phytoene synthase, conserved site (InterPro:IPR019845), Terpenoid synthase (InterPro:IPR008949), Farnesyl-diphosphate farnesyltransferase (InterPro:IPR006449), Squalene/phytoene synthase (InterPro:IPR002060); BEST Arabidopsis thaliana protein match is: squalene synthase 2 (TAIR:AT4G34650.1); Has 1469 Blast hits to 1468 proteins in 521 species: Archae - 28; Bacteria - 581; Metazoa - 114; Fungi - 174; Plants - 415; Viruses - 0; Other Eukaryotes - 157 (source: NCBI BLink).	MGSLGAILKHPDEFYPLLKLKMAVKEAEKQIPPEL HWGFCYSMLHKVSRSFALVIQQLGTELRNAVCV FYLVLRALDTVEDDTSIATDVKLPILRDFHKHIYDR DWHFACGTKHYKVLMDEFHHVSTAFLELDRGYQ VAIEEITRRMGAGMAKFICNEVETIGDYDEYCHYV AGLVGLGLSKLFHNTGLEDLASDDLSNSMGLFLQ KTNIIRDYLEDINEIPKCRMFWPREIWSKYVNKLE DLKYEENSEKAVQCLNDMVTNALFHVEDCLKYM SDLRDPAIFRFCAIPQIMSIGTLAICYNNLQVFRGV VKMRRGLTAIVIDQTNTMPDVYGAFYDFACMMK PKVDKRDPNATKTLSRIEAIQKICMDSGTLNKRKL HVVKNKSAYTPIMMVVFFIVLAVMFARLSANRQN NYLSAVMAFL
103	sp_191200_fjgi	0.978481	0.9443563	0.9607737	uncharacterized protein	MILLSVYCNLLSKFGILAEVSLSISQTHIRGVVCLVI LVPSKILVHITWLVNWRYFLKLFFVCVLAEVWYT NMKWLF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
104	sp_191910_jamq	0.959439	0.942882	0.949557	LATERAL ORGAN JUNCTION (LOJ); CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT3G54980.1); Has 63662 Blast hits to 14403 proteins in 306 species: Archae - 7; Bacteria - 57; Metazoa - 980; Fungi - 1296; Plants - 59204; Viruses - 0; Other Eukaryotes - 2118 (source: NCBI BLink).	MRNLSKSLQISSTLLRSSVTHKPLSSSLTSLSSSE PLISQDPISESSLPNQSPISNDFNEGYSSIPISQNH QNPTNFSHGINGLLSSKDPPSAMEYFKWAEKQP GFGEGVDSFCLLLHNFMGLEGQQCAVKELVSTQ NFPDAKTVIRGLNDCSNRFGFDLDSRIYNYLLCG FIKTGRIEDSISCVEWMLENKIDLWVSYLNDLIYEL GIQRKMIKEAKELYDKVFCKGNVVNGGTVRIMMK ACLKDGNPSEADRYFRLARDKGVELDPAAFSCAI VAVSQKPDSNAACELLMEMKEKGWVPSEGTYT QTIIACVKQRNLMEALRLKDEMVSCGKPMNLVVA TTLMKGYCVSGDLESALGLFAKLDEYELSPNNIT FSVLIDGCCINGNMEKASELYDQMIEKGIEPSVYN VNSMIKGFLKARLFEKAYEMFNKAAESGLVDDSS LNSLLSCLCKDGNMDEAQKFWGMALQCGIEPNL QNYNNLILGYCRKGNVAMAHSMYVEMIGKGLNP NVFTFTILINGHFKKGEVDKALNLFDEMSDQRIDC TDHTYNTMIGGLCKHGKTSLARNMLENFIKEANF VPSCMTYNSIVGGLIKEGDLDSALSVYNEMGENG VSPDICTYTIFIDAFRKSNNFDLALKMQNEMISKG LELDITAYGTLIDGCCKRRDMESACKLFNELCKS GLSPNIYIYNNMISGYRNLNNMEVALSFQKKMTE QGVLCDLETYTTLIDGLLKEDKLKSALEFYDDMV EKGIRPDKITYTVLVKGLCKKGKLENARNFFKEM EEKGVTPNSLIYNALVAGYFREGNLPEAFKLYDE MLERGLTPDKMTFDFLMNANFKPGYQSSRTSSD MPIFCKEALA
105	sp_194050_pwuc	0.970968	0.941858	0.966504	embryo sac development arrest 39 (EDA39); CONTAINS InterPro DOMAIN/s: IQ calmodulin-binding region (InterPro:IPR000048); BEST Arabidopsis thaliana protein match is: calmodulin-binding family protein (TAIR:AT2G26190.1); Has 379 Blast hits to 276 proteins in 60 species: Archae - 0; Bacteria - 22; Metazoa - 1; Fungi - 104; Plants - 246; Viruses - 0; Other Eukaryotes - 6 (source: NCBI BLink).	MGLSVSILISAWHEILSQKLFTLVSDGLLSSPVKR FPVSFRRTESFQKSTTDQSSPNRISETNHCQSG ETGAKKDHHHKKPENSTSKTNRISETNHCQSTE MGAKKEHPNKPEKIVLEKSPSFNTLVQEYDPTRT SSVVVVTGSNGLVHKPLPAITLPEPAVLFSPRPV SELDAAAVKVQKVYKSYRTRRNLADCAVVVEEL WWKALDFASLKRSSVSFFNNDKQETAVAKWAR AKTRVAKVGKGLSQNEKAQKLALRHWLEAIDPR HRYGHNLHIYYDIWFQSESSQPFFYWLDIGDGKE INLEKCPRTKLQQQCIKYLGPIERETYEVIVEDGK LTYKQNGELVNTVGECKWIFVLSTSRSLYIGQKK KGQFQHSSFLAGAATTAAGRLIADKGVLEAIWPY SGHYLPTEENFKEFISFLEDNHVDLTNVKKCAYN EEGSFKVPEDESSPEKTTDITDESIPAEEVKTKTE ESEPFSFARRLSCKWVTGNGPRIGCVRDYPMDL QSRALEQVNLSPRVTPGSFGPIPSPRPSPKVRVS PRLAYMGLPSPRTQMAAA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
106	sp_199780_ehzg	0.966674	0.972538	0.938733	alpha/beta-Hydrolases superfamily protein; FUNCTIONS IN: hydrolase activity; LOCATED IN: endomembrane system; EXPRESSED IN: shoot apex, hypocotyl, root, leaf; CONTAINS InterPro DOMAIN/s: Alpha/beta hydrolase fold-1 (InterPro:IPR000073); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT4G36610.1); Has 8797 Blast hits to 8792 proteins in 1517 species: Archae - 89; Bacteria - 6094; Metazoa - 449; Fungi - 67; Plants - 445; Viruses - 0; Other Eukaryotes - 1653 (source: NCBI BLink).	MVNIVSCYMSMLHKIMKLAGMHPKEVEIEPGTVI HFWVNNNNDNDDDDNNAKKKSRRRPALVLIHGF AGDGILTWQFQVLALRKKYSLYIPDLLFFGGSHT TSKAKITPRFHADCFAKGLRKLGVEEKCVVVGFS YGGFVGFQMAEHHPELVDSMVVTGSVLALSESI SSKSLERIGFSSWPDYLMPNSVEGVKVLLDIGSY SFPWMPKVFYKHYLE
107	sp_199960_garc	0.968884	0.963791	0.997238	Chloroplast-targeted copper chaperone protein; FUNCTIONS IN: metal ion binding; INVOLVED IN: metal ion transport;; LOCATED IN: cellular_component unknown; EXPRESSED IN: 12 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Heavy metal transport/detoxification protein (InterPro:IPR006121), Copper chaperone, chloroplast-targeted, predicted (InterPro:IPR016578); BEST Arabidopsis thaliana protein match is: Chloroplast-targeted copper chaperone protein (TAIR:AT3G53530.1); Has 1086 Blast hits to 1014 proteins in 44 species: Archae - 0; Bacteria - 10; Metazoa - 2; Fungi - 4; Plants - 1070; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MPIINIKNIKGIDFSCASPASTAICSSTDQNFMVKN TTKSLNRLNSYLNDHYKNTLNQHQSSSPRAPCIS ELPFTPRLSSSFREKTRKKSPSPSIVKHSDLTPRR KSTADIADVRRRYYDKCSNNSSILSPPHSHSSRY LLGESKLLDSLSESHQNVTNYRKSSSDALPLVPS TPLLVSREAHETKSFGSFLDVQKGNKSLTIVNPS DQRLIKSSKYDYDGHNQRHKSMGYDTNARNDSL VLKKSSSSAARSRDQVVVLRVSLHCKGCEGKLR KHLSKMEGVTSFSIDLETKVTVIGNVTPLSVLTSI SRVKNAQFWPSAASASPSSSSSSSSTSSTIDFN
108	sp_202730_niqq	0.945332	0.955677	0.974815	high mobility group B3 (HMGB3); FUNCTIONS IN: chromatin binding, structural constituent of chromatin, DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: chromatin assembly or disassembly; LOCATED IN: chromatin; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: High mobility group, superfamily (InterPro:IPR009071), High mobility group, HMG1/HMG2 (InterPro:IPR000910); BEST Arabidopsis thaliana protein match is: high mobility group B2 (TAIR:AT1G20693.2); Has 5742 Blast hits to 4932 proteins in 507 species: Archae - 0; Bacteria - 0; Metazoa - 4274; Fungi - 383; Plants - 553; Viruses - 0; Other Eukaryotes - 532 (source: NCBI BLink).	MKGGKGKAETTTTRRGGDAKLSVKKGAPAKKG KPAKDPNKPKRPPSAFFVFMEEFRVTYKQKHPN NKSVAVVGKAGGEKWKSLTDAEKAPYIQKAEKR KADYEKNMTTYNNKLAEGKDAEEEEESDKSRSE VNDEDDEDGSGEEDEDDD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
109	sp_213800_kcjw	0.9694426	0.941223	0.9377788	UDP-glucosyl transferase 74D1 (UGT74D1); FUNCTIONS IN: transferase activity, transferring hexosyl groups, UDP-glycosyltransferase activity, transferase activity, transferring glycosyl groups; INVOLVED IN: metabolic process; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT1G05675.1); Has 8199 Blast hits to 8111 proteins in 470 species: Archae - 0; Bacteria - 484; Metazoa - 2414; Fungi - 25; Plants - 5092; Viruses - 117; Other Eukaryotes - 67 (source: NCBI BLink).	MTNLWRVRTVGPLLPSSYLDKRIKSDVDYGINPY KPNNEACMNWLNTKQAASVVYVSFGSVANLSVE QIAEIAKALKKIPSSFLWIVREAEQEKLPEDFITET SEKGLVMSWCTQLDVLAHEAVGCFITHCGWN
110	sp_214020_fsci	0.9838695	0.9678058	0.9718378	uncharacterized protein	LSRALAVLASASLGSREREEFLRHVNKE
111	sp_214780_pcju	0.9599638	0.963224	0.9838695	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: basic helix-loop-helix (bHLH) DNA-binding superfamily protein (TAIR:AT2G22750.2); Has 2910 Blast hits to 2902 proteins in 174 species: Archae - 4; Bacteria - 2; Metazoa - 77; Fungi - 51; Plants - 2768; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MDDKEPTPEITAPDIKVRFVNDNVLITIVHEIPHGL VTKIMREISNLHLFIANMNSMPIGNISQVSIVAEVS PKPY
112	sp_000230_zzhd	#N/A	0.9447247	0.9838695	S-adenosylmethionine carrier 1 (SAMC1); CONTAINS InterPro DOMAIN/s: Mitochondrial carrier protein (InterPro:IPR002067), Mitochondrial substrate carrier (InterPro:IPR001993), Mitochondrial substrate/solute carrier (InterPro:IPR018108); BEST Arabidopsis thaliana protein match is: S-adenosylmethionine carrier 2 (TAIR:AT1G34065.1).	TAGAVGGMDASLIRVPTELVKQRMQTGQFVSAP DVVRLIVAKEGFKGLPIVSASRFAV
113	sp_000570_tqfy	#N/A	0.9623397	0.9878638	Peptidase family C54 protein; CONTAINS InterPro DOMAIN/s: Peptidase C54 (InterPro:IPR005078); BEST Arabidopsis thaliana protein match is: Peptidase family C54 protein (TAIR:AT3G59950.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	AHVPDQVKNSEVANIAKDDIEAQFALSIDPSLEIG FYCRDI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
114	sp_005800_dxse	#N/A	0.928442	0.920077	alpha/beta-hydrolases superfamily protein; FUNCTIONS IN: hydrolase activity; INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: stem, hypocotyl, root, flower; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Lipase, GDXG, active site (InterPro:IPR002168), Alpha/beta hydrolase fold-3 (InterPro:IPR013094); BEST Arabidopsis thaliana protein match is: carboxyesterase 13 (TAIR:AT3G48700.1); Has 8596 Blast hits to 8578 proteins in 1414 species: Archae - 104; Bacteria - 4918; Metazoa - 486; Fungi - 808; Plants - 1365; Viruses - 3; Other Eukaryotes - 912 (source: NCBI BLink).	MASLIENHNLLIVIFLLLPLLAQISSSETTQPPIILQP YIIDYQNGTVLRYNSTTPPTVVPPPNDAHVQSKDI TFPASSDRNLRARVFLPAHPDPHRKIPVLVYFHG GAFCIGAPPYSGDSNFVARVAAEVGVLAVAIDYR LYPEFTVPAPFEDAWASLQWIVAHASSTGPDHD PWLARYGDLQRIYVGGDSAGGTIAHNLAMKAGL RPNTRFIGLFLAMPYFLGSKRVPLEPEGIASSTNS KVWSYVCNNCTNGVDNGFINPFGPEAPSITHLG CKKLLVYTAEVDELRARGIWYYKMVKASKWAGQ AEWIEAKGETHVFHIIKPDEPATTIMQDLSTFINA D
115	sp_020020_fimf	#N/A	0.946663	0.971385	uncharacterized protein	RNARSMLLLAAEAHEADAEMILIPNQSPRKADDV
116	sp_021420_dzdt	#N/A	0.924322	0.968624	uncharacterized protein	TNIANVDTHHCFPIFSSELNEKRKNWRSPLLQIQA IAIERFSLEGARQWQLLDRLPPKPPFEIKHHEVAI VKVNNLPSSRVNLKEYLDRVFFYDAVMNEEASN DKGLRLKKLSQVNDIFIYFILLDE
117	sp_023630_ckhw	#N/A	0.944539	0.937849	Domain of unknown function (DUF220); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF220 (InterPro:IPR003863); BEST Arabidopsis thaliana protein match is: Domain of unknown function (DUF220) (TAIR:AT1G23560.1); Has 170 Blast hits to 157 proteins in 17 species: Archae - 0; Bacteria - 4; Metazoa - 0; Fungi - 0; Plants - 157; Viruses - 0; Other Eukaryotes - 9 (source: NCBI BLink).	MSSTSQSTDINHQRRQLIDPAFPKFFTQAPQRLQ NFFKSHLTGLRKAGQRKDVHELQQDIRKEQKSS TDLEADLQKQLQLWRANPHWGSPPPEINVTVPK GSLCNLNVEFDVGLLPPDAVYNIVIDPDSKRVFRNI KEVISRKVLVDEGSRQVVEVEQAAIWNFLWLSG TISVHVLVDQNREDHTMRFKQVNTGFMKKFEGC WRVEVPVFVDEKMCSPFKPKTLSDYRSCTGGKG RIGSRVSLEQLLQPAIVPPPPISWYLRGITSKTSE TLMNYLLEEVARIKGISQVAGSNGDEKSVSEITNE PDIDEVRDIKKRWALRRRHALHYRKK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
118	sp_025850_ngzk	#N/A	0.5277034	0.9428837	Pectin lyase-like superfamily protein; FUNCTIONS IN: polygalacturonase activity; INVOLVED IN: carbohydrate metabolic process; EXPRESSED IN: stem, root, stamen, seed; EXPRESSED DURING: 4 anthesis, E expanded cotyledon stage; CONTAINS InterPro DOMAIN/s: Pectin lyase fold/virulence factor (InterPro:IPR011050), Glycoside hydrolase, family 28 (InterPro:IPR000743), Pectin lyase fold (InterPro:IPR012334), Parallel beta-helix repeat (InterPro:IPR006626); BEST Arabidopsis thaliana protein match is: Pectin lyase-like superfamily protein (TAIR:AT4G01890.1); Has 5413 Blast hits to 5130 proteins in 851 species: Archae - 6; Bacteria - 1539; Metazoa - 261; Fungi - 1343; Plants - 1850; Viruses - 100; Other Eukaryotes - 314 (source: NCBI BLink).	MPLHVQPNDQYSQISIPPSYSPNIFDSPSDDDHS VGYNSPAPPPSDDDDHGVSGSTDAHRRRRQTH HKPSGHKAAGHRHQAAAPAPSHRQPAGPTGVF DVRSFGAVGDGLVDDTLAFKSAWDAACGGDNA KVVAPKGFSFLVQPLIFNGPCKSGVVFQLEGTIVA PDGPDAWPKDSSKRQWLVFYRIHSLTLQGGGLI DGRGDKWWNLPCKPHKGINGTSTHGPCDSPVA MKFSGNTNLTVRGLFLINSPQFNIRIDGCQDVNV DSLIIKSPANSPNTDGIHVENTNDVTIYNSIIANGD DCISIGAGSHNLDIRNVTCGPGHGISIGSLGIHQT KACVSNITVSNCIIKHSDNGIRIKTWQGGSGSVSG INFEGIHMDTVRNPIIIDQYYCDSAKCSNQTSAVSI NDITYSNIKGTYDTRSPPVHLACSDPAPCTNITIS GLELLPAPVTGMDLVTDHKHQLVSMEPFCYNAY GNNIGEVTNPPVYCLLDGMPYTSSDVEQC
119	sp_031470_htwz	#N/A	0.9641122	0.8337703	Eukaryotic aspartyl protease family protein; FUNCTIONS IN: aspartic-type endopeptidase activity; INVOLVED IN: response to salt stress; LOCATED IN: cell wall, plasma membrane, membrane, plant-type cell wall; EXPRESSED IN: 6 plant structures; CONTAINS InterPro DOMAIN/s: Peptidase aspartic (InterPro:IPR021109), Peptidase aspartic, catalytic (InterPro:IPR009007), Peptidase A1 (InterPro:IPR001461); BEST Arabidopsis thaliana protein match is: Eukaryotic aspartyl protease family protein (TAIR:AT1G03230.1); Has 1398 Blast hits to 1391 proteins in 46 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1396; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MASYLNHHPFIFYLLIIYTISISISISISISTTSALFL PKALVLTVVKDPSTLQYTTKLIQRTPQPLSLDLGT PISWVICDNRYISSTYHPAYCGSPQCTLANSKLC RTCPPPAKPGCNNNTCTTTSLNNPFTNIIISPPGG TELATDIVRIPTTDGSNPGQLVTVPNHLFVCAPKS LLQGLAKGVSGVAGLGRSKISLPSQVSSYFRFPR KFTLCLPPLTESGVVFFGDVTYNFNPPGGAASAN LSYTLLVANPFRPSDYFIGVQGIKVGGEDVPSVN KSKLSINSKGNGGTMISLSHPYTIMETSIYKAITTIF NEQLTTVVKRVNPVAPFQFCYDSNTIGFSPIGPM VQDIDLVLQSKEVKWTIAPLNSMVRMNDKAYCLG ILDGGNNPSTSIIIGGFQIEDNLLQFDLDKSMIGFG SSLIYNRMPCSNFNFTSNVI
120	sp_044330_oxia	#N/A	0.9538268	0.8377231	Arabidopsis protein of unknown function (DUF241); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF241, plant (InterPro:IPR004320); BEST Arabidopsis thaliana protein match is: Arabidopsis protein of unknown function (DUF241) (TAIR:AT2G17070.1); Has 750 Blast hits to 734 proteins in 87 species: Archae - 2; Bacteria - 37; Metazoa - 79; Fungi - 28; Plants - 534; Viruses - 2; Other Eukaryotes - 68 (source: NCBI BLink).	MASYHTRSFSFPSNSHPVADQLDAQLSRLRSSQ AASTSPLTSKLSGLNDLYKCVEEFLQLPQNQKTV SQSQGKNVVEQVLDGSLRLLDIVSTSRDVLALSK ERLQDIQSVLRRRCSGEIDITNEVSEYLKTRKASK KIIKKCLKDINAVDKTNESIXXGEIDITNEVSEYLKT RKASKKIIKKCLKDINAVDRTNESIAIESMLKDVQT LTVDVFKSLLSYIGGSQKSSWSFSKLVRQGSDKE EAASISVFEAVDATLSQKNKANVDMSQLVKLESEI QEIDEVLESLFRHLIKTRATLLNVLSN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
121	sp_044550_qdkt	#N/A	0.9295434	0.9288636	Kunitz family trypsin and protease inhibitor protein; FUNCTIONS IN: endopeptidase inhibitor activity; INVOLVED IN: biological_process unknown; LOCATED IN: apoplast, cell wall; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Proteinase inhibitor I3, Kunitz legume (InterPro:IPR002160), Kunitz inhibitor ST1-like (InterPro:IPR011065); BEST Arabidopsis thaliana protein match is: kunitz trypsin inhibitor 1 (TAIR:AT1G73260.1); Has 1019 Blast hits to 1019 proteins in 106 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1018; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	MAHHLLLPAATLLLTLLFLSPPTATAKDSVVLDVD GNPVEAGSQYYVLTSGWGGAGRGGLTAASLGG PCPLYVGQYRLWVQKGGSPVTFYPSNPSQKHITQ SSDVNIAFAPNPICRNQGVWQLGYNGISFYITTN GVIGKQEDVDTFANWFKIEQAGGLPGAYKIVYCF AVPSTGPEPVVPDSRRRTGPEPDTEYCEQLDAT VPVFPMGLSLLTPVPIDNPLPFFGFVFKKVNTTAN HSY
122	sp_062500_hayx	#N/A	0.9332621	0.9344593	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: lipid glycosylation, biosynthetic process, carbohydrate metabolic process; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 28 (InterPro:IPR004276), Glycosyl transferase, family 28, C-terminal (InterPro:IPR007235); Has 7866 Blast hits to 7866 proteins in 2456 species: Archae - 0; Bacteria - 5567; Metazoa - 4; Fungi - 2; Plants - 34; Viruses - 0; Other Eukaryotes - 2259 (source: NCBI BLink).	MVVVATISQLSLQFPPKHTQFKPIYPPKPRSQVF CCLSIDPPTLESSNVEERKGKQLRVVFAAGGAG GQIFPAVAIADEIRIINPDAQILFIGTDAGMESTAVP ASGYEFSAIPATPLSRPILSVNNLSIPLKWVKSMA ASWKIMKDFKPEIVVGTGGYVSFPICLIAAIKGLR MVIQEQSSIPGLANWFLSLVADLVFVAYNSSVDR FPTEKKKIVVSGNPVRLLIKKFVSKAVARTYFFPR LAKVSESEAVVLLVLGGSFGANSMNIALFNIYYFQ MLEQRPNLHIIWETGVESFNEMESLVRNHPRLVL APFLRSMDMGYGAADLIVSRAGSMTCTELLVTG KPAILIPSPNDAEGHQFQNASLMADLAGSRILTED ELDSTTLRNAITDIIDNDVLMSAMSDRALQAAKPN AGEEIAERVIALVESSSVKA
123	sp_062670_phih	#N/A	0.9433208	0.9333361	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G34560.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MNPLNINHIKVEKAKAIKLRYQKLRKITRFFRVFEII LILAVISRLSLQLPLALFKLSNECFYSLSIFIVSPKF VFIIGNAIVITLFVNSGRLSSNECSPKTTKTPPYQR KNKNMMKVDQKSCKKQGIITEENAMMAMIFPPS EQQQQVIKIYQRSKSEKMIIKREKVEKKCCGLRR SNTEIKSLKGCGCSNDVKECHYPEDHMSNEEFQ KTIENFIARQQKFLREEEFSALKE
124	sp_063080_tirx	#N/A	0.9633832	0.9333359	uncharacterized protein	MASYEASQQVTQKFMEGNTSRRRMSAAMVATM EMRLELISFGELDGRTNGGGG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
125	sp_073260_kiyo	#N/A	0.945616	0.949964	Pentatricopeptide repeat (PPR) superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885), Tetratricopeptide-like helical (InterPro:IPR011990); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT4G39530.1); Has 56701 Blast hits to 14050 proteins in 283 species: Archae - 2; Bacteria - 22; Metazoa - 148; Fungi - 141; Plants - 55528; Viruses - 0; Other Eukaryotes - 860 (source: NCBI BLink).	MKRFLFSTPNIHSLSSIKHLIAQNLHPQALKNSLFL SSPSLTDQTYALFLKSGHVLDTFLATSIITHFSKS GDFSRSNGFLLDAHGVDTVAFNAMLSGYAQFHQ SSPSFALYNKLTCYGLSPDKYTLSSLIKSCDNLEK IRVSHVICVKYGFNLNGFVVSGLIHKYGKGGDVE SAEKCFEECLDLDSVVWTAMINGYVWNGEFEKG RQIFMAMRGLGLELNEFSLTAVVGALREVREGE QIHGFSFKFGLIFDGSLHLSNVVMNMYSKCGHRL DALKVFDEIPNPDVVSWTSRIGASRDGGEALEAL KFMLLNNVAVNELTVVNVLTIIEEPRLLIAGRQVH SLCHKAGYLTVLSVNNALISMYGKCREMDDARLV FDELDFYDPISWNSLIGGYSDNGLLSAVLCIFHQ MRDLGVQHDLYTLASVLGAVTGPQFVDLVMQIH SHIIKSGLVSDDSMISCLISSYGKCSCLDHAKKVL SSTQEVNIGHLQVMAATAVTAGYPADAVECFKS AMNLWSEMNGTTLSLIFKACSSLTNLGQGKAIHC LALKAGLGSDRFVETAIVDFYCKCGSIADARNIFV DVSSHNLAAWNAMIMGFAQHGCYNEVLSLFQR MTELGIKPDEITYLGILQSCCHAGLVNEANIHLNS MSELHGLLPCLEHYATMIDLLGRVGLLKEARLVIE KLPVTPDAQVWQSLLSACHNSGNLEIGEVAAKEL LKLQPENDSAYILLSNLYARAHDWDAVRRLRRE MKDKFIYKEPGSSWIEAGGSVYHFLAEDTSHPEN EKIYFYLQSIAQQMLPLSEVTEDSFVCMDGW
126	sp_078810_rqrq	#N/A	0.92352	0.931959	uncharacterized protein	PAAQHFNDLYKAVQSSMDLYYVYRWNQLQLTQ NSIDMAVLRKRAEDVESLLEDSKKKLDAAASDLE AANKRLSDIEPKLQAEAERADGLDQQLASINASL PGVKKSAAKKAVDKLLQSDWFNDILTLRHNGGW CAAHRVVCHVKKLDEDGWQVFEDGYEEKELYK
127	sp_113110_jrhq	#N/A	0.923957	0.931881	HSP20-like chaperones superfamily protein; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20 (InterPro:IPR002068), HSP20-like chaperone (InterPro:IPR008978); BEST Arabidopsis thaliana protein match is: HSP20-like chaperones superfamily protein (TAIR:AT2G27140.1); Has 638 Blast hits to 638 proteins in 68 species: Archae - 0; Bacteria - 6; Metazoa - 0; Fungi - 34; Plants - 583; Viruses - 0; Other Eukaryotes - 15 (source: NCBI BLink).	MAIKTKSSTPGSPSYVDFHPACDFLKAEGIETLIIH LPDFKKEQLRVQVNKEGVVIVSGERATNEDGSK RSRFVKETKIPEGCDINDIRAKFINGRLNVTMPKKI LITPPPPPPPPPPQLIQEQPPSSSSPPPPPPKPEA VTDQKPKPEAASVSDDQTPKGSTIIGGTETSCSY HGQPKSINRMFELGSVIQSRLQGRKNVALGFGIA VVTMVAIGAFVASKLGSDTTSSPSLYIEDSGLM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
128	sp_116810_wrcx	#N/A	0.9603848	0.9633631	annexin 4 (ANNAT4); FUNCTIONS IN: calcium-dependent phospholipid binding, calcium ion binding; INVOLVED IN: in 6 processes; LOCATED IN: cell surface; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Annexin like protein (InterPro:IPR015472), Annexin repeat (InterPro:IPR018502), Annexin (InterPro:IPR001464); BEST Arabidopsis thaliana protein match is: annexin 8 (TAIR:AT5G12380.1); Has 2371 Blast hits to 2094 proteins in 191 species: Archae - 0; Bacteria - 0; Metazoa - 1728; Fungi - 90; Plants - 415; Viruses - 0; Other Eukaryotes - 138 (source: NCBI BLink).	MADVSHELQAVTKAFSGLGVDEKSLISIAGKWQH DHKNSFRKASPSHFLADKERHFEKWNEQQVELI KHEALRFKSALVLWTMHPWERDARLLKEALIKG PEGYNLVIEIACSRSSEELLGARKAYHSLFDSSVE EDVASHVDGHKRKLLALVSAYRYEGSRVHDGN AKSEAEVLHRAINNNISDDDEELVRILCTRSKLHIS AVFKHYKQLFGKDLSEEFHGNLSLKAALECLVIPE VYFSEVLDAAMREGATDFTQEGLTRVIVSQADG NMEEIKAAYHNLFGAHLSHKIEETCLGNFKDFLLT LVARGE
129	sp_120600_qjpz	#N/A	0.9221335	0.9583121	hydroperoxide lyase 1 (HPL1); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, heme binding; INVOLVED IN: fatty acid metabolic process, response to wounding; LOCATED IN: chloroplast envelope; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group IV (InterPro:IPR002403); BEST Arabidopsis thaliana protein match is: allene oxide synthase (TAIR:AT5G42650.1); Has 6712 Blast hits to 6702 proteins in 615 species: Archae - 25; Bacteria - 392; Metazoa - 3265; Fungi - 548; Plants - 2246; Viruses - 0; Other Eukaryotes - 236 (source: NCBI BLink).	MMMTISATMSSTPNPIPTTLTFPSPPPPANLPLR PVPGSYGLPVLGPIGDRLNYFWFQGQDSFFKKK MEKYKSSVFRTNVPPCFPFFAGVNPNVIALLDVP SFSYLFDMDIIDKKNVLVGDFMPSTKFTGDMRVG VYLDTTEEKHTKIKNFSTDILKKSSGIWVSSLLSN TEIMWETVEESLNKDGSTKIFTPLQKCIFNFLSQV MLRANPSNYSSELCENGHIMIDKWLAIQLLPIINN MFQPLEEIFLHSWSYPYFLVKGDYEKLVNFVAKE AREVVETGVTEFGLSEQEAIHNLLFMLGFNAFGG FSRFLPALVNNLRENKGGIQDILRKEVRGKCNSL TSLTFKSVQDMPIMTSFVYETLRLQPPVPLQFGR ARKDFELRTHESRFEIKKGELLCGYQTLVMRDPK IFDDPDTFVSERFMGETGGELLNYLFWSNGPQT GKSDASNKQCAASDYVPLTACLFLAHLFLRYDSI TLDSSGAITKVQKAT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
130	sp_131330_pwyu	#N/A	0.8837217	0.8228631	disproportionating enzyme 2 (DPE2); FUNCTIONS IN: 4-alpha-glucanotransferase activity, heteroglycan binding; INVOLVED IN: polysaccharide metabolic process, circadian rhythm, maltose catabolic process, maltose metabolic process, starch catabolic process; LOCATED IN: cytosol; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Immunoglobulin-like fold (InterPro:IPR013783), Carbohydrate-binding-like fold (InterPro:IPR013784), Glycoside hydrolase, family 77 (InterPro:IPR003385), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, carbohydrate-binding (InterPro:IPR002044), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: disproportionating enzyme (TAIR:AT5G64860.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MVNLGLFSSENGVKSVSVNFRIPYFTQWGQSLL VCGSVPSLGSWNVRRGLLLSPTSQGSELIWGGT IEVSSGCECEYSYYVVDDDKNILRWEMGKKRKLL LPSSIQDGEIVAFHDLWQAGSDTLPSSSAFKDVIF RNSSSLTVERPLGVLQNKLGLEDSVLVNFKICCP NLDEGTSVYVIGSPQKLGQWKTEGGLKLSYAGD SVWQADCVIGKEDFPLKYRYGKYDKAGKFAIEIG TNRKLLVESLSNQARYIFLSDGMMREMPWRGSG MAIPMFSVRSEEDLGVGEFLDLKLVDWAVASG FHLVQLLPINDTSVHGMWWDSYPYSSLSVFALH PLYLRVQALSENIPEEIKLEIVKAKERLDVKDVDYE ATMATKLSIAKKIFSLEKDLTLKSSSFQKFFLENEE WLKPYAAFCFLRDFFETSDHSQWGRFSDFSKEK LEKLVSDDSPHYNIVCFHYYIQYHLHLQLSEAAEH ARKKGVVLKGDLPIGVDRNSVDTWVYPNLFRMN TSTGAPPDYFDKNGQNWGFPTYNWEEEMSKDNY GWWRGRLTQMAKYFTAFRIDHILGFFRIWELPE HAVTGLVGKFRPSIPLSQEELEKEGIWDFDRLSR PYIRQEFLEEKFGSSWTIVAALFLNECQKGHYEF KEDCNTEKKIASKLKDSTGKSLLLGSEAELRINLF DLLRNIVLIRDPEDTSKFYPRFNLEDTTSFTDLDD HSKNVLRRLYYDYYFHRQENLWRQNAMKTLPVL LNSSDMLACGEDLGMIPSCVHPVMQELGLIGLRI QRMPNEPDLEFGIPSQYGYMTVCAPSCHDCSTM RAWWEEDEERRRRFCKTVMGSNKLPPSKCIPEI AYFIIRQHVESPSMWAIFPLQDLLALKNEYTTRPA TEETINDPTNPKHYWRFRSHVTLEMLLKDQELKT AIKELVRSSGRSYPHDDDETVAVSNKGNEPPAV ASKIVQPNGASQEEEEDAVAVAVL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
131	sp_133890_wxdc	#N/A	0.9333336	0.9333336	disease resistance family protein / LRR family protein; INVOLVED IN: signal transduction, defense response to fungus, defense response; LOCATED IN: cell wall; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat transmembrane protein kinase (TAIR:AT4G20140.1); Has 135121 Blast hits to 32840 proteins in 1181 species: Archae - 56; Bacteria - 8201; Metazoa - 34800; Fungi - 1569; Plants - 79859; Viruses - 2; Other Eukaryotes - 10634 (source: NCBI BLink).	MYSNSSHLLKLGCKPAERKALLDFKSSVTFLDKL VSSSWEGEDCCSWSRVKCNPTTGHIVELRLNG GNYSYLFEARKMESSLVYLLELNHLEYLDVSANN FNSAHDYFVNDFNLSKIPKFMGSMKQLRDCAFT GIEVGNIWSLVRNLCNLKSLDLSSNKIQAKLLEGR ENISRTICTHFDIEMLDLSWNSISGRIPVWLGEEF RSLKSLSLSGNNLSGEIPTSLGMLSSMKHLDISN NPLEAAVLSESHFANLSSLTDLFMDQTFLTLKLSS YWLPPFQLQNFYASYCKINGPFPQWLKTQKNIS NLYLSDAGIHGALPKWFHNMHSLSYVSLSNNQIT GCPIFPVNFEQISLSNNSLSCQYLTNMSRNTVYY SSQGVDLSDNSISGQLLEHLYHDMPNLSTLVLSR NLINGSIPESMCHFTSLQILDLHKNRLSGTIPHCL GESSELEFVHLAFNNLSGDIPCFRLTHVENEWSS SYLSFVSLNDNMLSGQIPSCLANLSTLEVLDVGE NKLSGEIPMFITGLNLPVLRILRLRNNKLEGSIPEQ LCSLSNLQILDLGLNYLGGTIPSCLANMTAMNSFV DYDSIFYADMNQVNEVIKGIERAYTTTLKYMVNV DLSCNNLIGSIPEGITNLTYLRNLNLSYNHLSGQIP MSIGCLKSLESLSLDLSRNKLQGTIPISMGAIFSLESL DLSYNNLSGQIPTGNQLQTLNDPLSYAGNPYLCG DPLAKKCKSDNDSPNGSSGDQNNEDEVSEDKH ETMWFYLVVMSGFATGFWCVVGTLFLKKSWRH AFFRRVEFVQDWLYVAVVVRVPKLMRRFI
132	sp_149180_nwmy	#N/A	0.9331481	0.9340272	acyl-CoA sterol acyl transferase 1 (ASAT1); FUNCTIONS IN: acyltransferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Wax synthase (InterPro:IPR017088); BEST Arabidopsis thaliana protein match is: MBOAT (membrane bound O-acyl transferase) family protein (TAIR:AT5G55350.1); Has 801 Blast hits to 791 proteins in 235 species: Archae - 0; Bacteria - 473; Metazoa - 0; Fungi - 36; Plants - 238; Viruses - 0; Other Eukaryotes - 54 (source: NCBI BLink).	MVTRLRPTVYLPTLEYSTKVVGRTWATLPAVMS TFFVSAIMHELIFYYLGRNWPTFEVTWFFLLHGL CLCVEIVAKKLVGGKWRIPRWISGPATVLFVVGT GFWLFLPPLLKAGLDTRPFQEFAAVAKFVRSLKA ALTF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
133	sp_155800_psps	#N/A	0.924565	0.926107	glutamate receptor 2.3 (GLR2.3); FUNCTIONS IN: intracellular ligand-gated ion channel activity; INVOLVED IN: cellular calcium ion homeostasis, response to light stimulus; LOCATED IN: endomembrane system, membrane; CONTAINS InterPro DOMAIN/s: Extracellular solute-binding protein, family 3 (InterPro:IPR001638), Ionotropic glutamate receptor (InterPro:IPR001320), Extracellular ligand-binding receptor (InterPro:IPR001828), Glutamate receptor-related (InterPro:IPR015683), Ionotropic glutamate-like receptor, plant (InterPro:IPR017103); BEST Arabidopsis thaliana protein match is: glutamate receptor 2.2 (TAIR:AT2G24720.1); Has 5318 Blast hits to 5220 proteins in 711 species: Archae - 48; Bacteria - 1305; Metazoa - 3164; Fungi - 0; Plants - 641; Viruses - 0; Other Eukaryotes - 160 (source: NCBI BLink).	MMKKQCCIFYVMMMNVFVFIPILVLGQGNTTETT PRELRIAIPVREFKDWVSVEHDPNTGEQIISGFAI EVFKAVVNKGLLSPIPYKFIPFVKSDGVTMNGTFD DMLRAVFNQ
134	sp_166240_sfrz	#N/A	0.921997	0.974644	O-methyltransferase 1 (OMT1); FUNCTIONS IN: myricetin 3'-O-methyltransferase activity, quercetin 3-O-methyltransferase activity, caffeate O-methyltransferase activity; INVOLVED IN: lignin biosynthetic process, flavonol biosynthetic process; LOCATED IN: cytosol, nucleus, plasma membrane, cytoplasm; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Winged helix-turn-helix transcription repressor DNA-binding (InterPro:IPR011991), Plant methyltransferase dimerisation (InterPro:IPR012967), O-methyltransferase, family 2 (InterPro:IPR001077), O-methyltransferase, COMT, eukaryota (InterPro:IPR016461); BEST Arabidopsis thaliana protein match is: O-methyltransferase family protein (TAIR:AT1G77520.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSSIIGSNESQIHPTTEEEASCMAMCLATWSVPS MVFKTVLELNVLDIIYFTGPEAQLSAEEIVAKLPAT NPHAAVVLDRMLRLLASFTILTCTQQVGDDGQPK RLYGLGPVCQFLAKNDDEVSYAAFANLMHDKVI MDTWKHLKEAVLVGGSAFRNSNGTSFFDYLGTD SRFYKVFYEGMDHHSTIVMKKLLDTYQGFKGISTI IDVGGATGLTLKMILDKYPSIKAINFDLPHIIKDAPS YPGVEHVGGDMYDSVPKGDAIFMKWTFQSGTD EKCVKLLKNCHEALPEDGKVIICEYVVPELPKTSY QAHTAFIWDSIMLTFTGGKSRTKKEFEALGKEAG FEDFQLVCFACDTWVMEYLKKKKSDN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
135	sp_178980_gxfz	#N/A	0.96082	0.939271	alpha/beta-Hydrolases superfamily protein; FUNCTIONS IN: hydrolase activity; INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: stem, hypocotyl, root, flower; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Lipase, GDXG, active site (InterPro:IPR002168), Alpha/beta hydrolase fold-3 (InterPro:IPR013094); BEST Arabidopsis thaliana protein match is: carboxyesterase 13 (TAIR:AT3G48700.1); Has 8596 Blast hits to 8578 proteins in 1414 species: Archae - 104; Bacteria - 4918; Metazoa - 486; Fungi - 808; Plants - 1365; Viruses - 3; Other Eukaryotes - 912 (source: NCBI BLink).	MNSRAKEVKVELPGFIRVYEDDSVERFYGSPQV PPTLEDPETGVSSKDVIISADKNISARIYLPKIITNN HPILVYFHGGDFCVESAFSLHHQRYMNILASQAG VLVISVEYRLAPEHPLPIAYEDCWTALNWVTSLD DPWLTKHGNLDALFIGGDSSGANIVHNLAMKAG TQSLDHGVKILGAFFSHSYFLDEPSSTRLDVQMA YKIWDFVYPNALGGSNSPMINPMGDGAPSLAGL GCSKVLVVVAELEGQRNSEIRYYNAVKDSGFQG SVELFETKGEDHCFHIFNPDTHNSKLLFQRLTSFL RN
136	sp_180410_nknp	#N/A	0.924565	0.926107	uncharacterized protein	MQAHCKRVIKSLIHTVDNKNKTEKTRLLVHGKIIQ VVPRAAATAGSRGRSNCMQLYPYCCHREAVFQ EANHPGEEGLQLHLMGIVTVIDRCCQCNSSTLSN YDDGHHQQKTQIFSWITLIENCHHHHPHPHCCSA EQENSIPKNRGLQQRIAIVTYDHPGHAFRQKADL TKIFEIVL
137	sp_191450_ttik	#N/A	0.966747	0.96123	Low temperature and salt responsive protein family; INVOLVED IN: response to cold, hyperosmotic salinity response; LOCATED IN: endomembrane system, integral to membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised protein family UPF0057 (InterPro:IPR000612); BEST Arabidopsis thaliana protein match is: Low temperature and salt responsive protein family (TAIR:AT2G24040.1); Has 1255 Blast hits to 1255 proteins in 413 species: Archae - 0; Bacteria - 524; Metazoa - 40; Fungi - 292; Plants - 377; Viruses - 0; Other Eukaryotes - 22 (source: NCBI BLink).	MAEANRCHIAFEILLAILLPPLGCEFLICVVLTLLGY IPGIIYAIYVIAFRSSPSVGAEDHYRALPQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
138	sp_192500_nzmu	#N/A	0.972216	0.9931387	wall-associated kinase 2 (WAK2); FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, calcium ion binding, ATP binding; INVOLVED IN: cellular water homeostasis, protein amino acid phosphorylation, oligosaccharide metabolic process, unidimensional cell growth, response to salicylic acid stimulus; LOCATED IN: plasma membrane; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: EGF-like calcium-binding (InterPro:IPR001881), EGF-like, type 3 (InterPro:IPR000742), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), EGF-like region, conserved site (InterPro:IPR013032), Protein kinase, catalytic domain (InterPro:IPR000719), EGF-like calcium-binding, conserved site (InterPro:IPR018097), EGF-type aspartate/asparagine hydroxylation site (InterPro:IPR000152), EGF calcium-binding (InterPro:IPR013091), EGF-like (InterPro:IPR006210); BEST Arabidopsis thaliana protein match is: wall associated kinase 5 (TAIR:AT1G21230.1); Has 140654 Blast hits to 123720 proteins in 4781 species: Archae - 129; Bacteria - 13911; Metazoa - 65181; Fungi - 9585; Plants - 33448; Viruses - 437; Other Eukaryotes - 17963 (source: NCBI BLink).	MFSYLFQFITIALCTWIHPLLTSASSFQPTSDPPIS KPGCQDICGNVTIPYPFGVGSNCYHTLPYAVTCN TSLSHPRLFLNKYNLEIEQINIHTKFHLDQNTITVN TPLQRICNTRDANISSIDLSGTPYSFSSDNSVFVG GCGSSALVLDRNNNLGGCTSVCNNSHMITDDCI GRGCCRFDLPSSIEFYRVNIFNFGGNCTNAMLVN MSSPIISKAVDPSLFEVAVVPTVLEWTPGNITNPG DDPNIDCQGSADGYDCYCNYGTSGNPYLPYGC QVPSECKECKLGCIQPENHYVCVRNNSEDKAYN RLVLLVLLGGSIVGLMLCTGNISCIYLCIKRRKHMK MKEKFFQQNGGFLLKQQISLNQSGNNDTTKIFTA QELREATNNYSKQNILGQGGFGTVYKGILPNDQV VAIKKSMVSDQSQVEQFINEVVILSQTNHRNVVK LLGCCLEIQVPLLVYEFVSNGTLFEHIHSDDARWL SWENGLRISTEAANAIAYLHSAASMPIIHRDIKSSN ILLDDNFTAKIADFGVSRLVPIDQTQVTTLVQGTF GYLDPEYFLTSRLSEKSDVYSFGVVLVELLTKRK PISWETGMEDRNLAMIFISSVKNNQFFDILEPQLV KEASEEQLISIGNLVEKCLSVKGEDRPSMKDVAR VLEEVANQSNNEIHEMSSEQRPVQPAIEQDSSR QFKF	
139	sp_195330_fxin	#N/A	0.946982	0.9938625	THO5; CONTAINS InterPro DOMAIN/s: THO complex, subunit 5 (InterPro:IPR019163); BEST Arabidopsis thaliana protein match is: THO complex, subunit 5 (TAIR:AT1G45233.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MRPTTIPTATAQLGVPSPVGYAPAFGSAPGFGT RGFWFPPPSTLVSPVLVSDPITPGPSSAVTSYVL GFMNPSSSDCSVPTKSMTTLQTEVQSPAQKPGV SPLRSSRKSRTDMEKELVAFIKECQGNSKDEGP ELIEFDGLCLSKYQMYRMVAMDEYAGIKYVKVAS KIYTSRWKAELGAGGRGRSVMLEPGFG
140	sp_205140_yqdr	#N/A	0.9603338	0.9938641	uncharacterized protein	MSPLPSPFEQPHHHRTAPLFCLPFSPVTADAVVV INLWGGFTSSLPSRLLLSLVPAM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
141	sp_207110_fdft	#N/A	0.944165	0.92315	S-locus lectin protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, sugar binding, protein kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation, recognition of pollen; LOCATED IN: endomembrane system; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480), Protein kinase, ATP binding site (InterPro:IPR017441), Apple-like (InterPro:IPR003609), PAN-1 domain (InterPro:IPR003014), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), S-locus glycoprotein (InterPro:IPR000858), EGF-like (InterPro:IPR006210); BEST Arabidopsis thaliana protein match is: receptor-like protein kinase 4 (TAIR:AT4G00340.1); Has 116366 Blast hits to 114909 proteins in 4332 species: Archae - 99; Bacteria - 12770; Metazoa - 42761; Fungi - 9470; Plants - 34242; Viruses - 402; Other Eukaryotes - 16622 (source: NCBI BLink).	MLDDMGTFSASDQLKFSALDVGFGIKRRLTMDF DGNLRIYSLNKLSGLWNITWTAVAKPCDIHGLCG RNGICEYTPEPKCSCPPSYEPTDISDWSKGCRP RFHRSCSDSEFIEMSHVDYYGYDLNSTTPASYDD CREICLGDCRCQAFNYRLTGEGMCFTKSALFNG VKSVSFPASISNQRFKWVYLYSFAIVIGTVEVILLI AGWWFLFGKHSLSGSLEDGYRAISNQFRSFSYN ELKTATRNFKEVLGRGGFGAVYKGVLADEREIAV KKLENVVQGEEEFWAEVSTIGKINHMNLARMWG FCSGRKHKLLVYELVENGSLDRHLFSTKTVLGW KERGGHGSSELTRIRGTKGYMAPEWAINLPITAK VDVYSYGVVILELIKGIRLSSWVVDDQNSSEEIPE LVRFVRLAKRKIQNGEDSWVEDLVDPRLEGKFS RNQAAMMIEVGLSCLDDDRNKRPTMESVAQALA ECEDEIL
142	sp_207690_ppmt	#N/A	0.929736	0.937088	uncharacterized protein	MDEYAGIEYVKVASKMYTSRWKAELGAGRGRSV MLEPGFGVKVQTNEPPQSLVAPFGAPFENLVIRE MFVVVVPVLHMYSTLPPHRRWYCVAISSMKDKRI WVLDPTSDEPLSEHGRLISHMLKYLDTLFYSKDE SWVKDGLSGWGVDLVSVPPSDDVSSCVVMLA WIKDIVQRNNLGK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
143	sp_211990_qsxc	#N/A	0.944114	0.9351185	ANTHOCYANIN11 (ATAN11); CONTAINS InterPro DOMAIN/s: WD40 repeat-like-containing domain (InterPro:IPR011046), WD40 repeat 2 (InterPro:IPR019782), WD40 repeat, conserved site (InterPro:IPR019775), WD40-repeat-containing domain (InterPro:IPR017986), WD40 repeat (InterPro:IPR001680), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT3G26640.1); Has 4731 Blast hits to 4446 proteins in 351 species: Archae - 6; Bacteria - 291; Metazoa - 1848; Fungi - 1200; Plants - 767; Viruses - 0; Other Eukaryotes - 619 (source: NCBI BLink).	MGASSEPSMNPEPFSDEQHKQRSEIVTYEAPWH IYSMNWSVRRDKNYRLAIASLLEHAQNRVEIVQL HDSTGEIKSDPSLSFDHPYPPTKAIFIPDKDCSKP DLLATSSDYLRIWQINSDSNSSSKIELKSLLNGNK NSEFCGPLTSFDWNEAEPRRIGTSSIDTTCTIWDI ERETVDTQLIAHDKEVYDIAWGGAGVFASVSAD GSVRVFDLRDKEHSTIIYESSEPDTPLVRLGWNK QDPRYMATIIMDSPKVVVLDIRFPTLPVVELQRHQ ASANAIAWAPHSSCHICTAGDDMQALIWDLSSM GQPVEGGLDPILAYTAGAEIEQLQWSSSQPDWV AIAFSNKLQILRV
144	sp_001340_gcdp	0.9333673	#N/A	0.9338251	uncharacterized protein	MTDSLPRFIVLGRPTGGVTYLRYVVEKGAQHRRL HLSSPEVISDRNFTKLEVVRAPPTIGLDMVHIRSC YNNKFLVLQSEGSNVFVAEADEPEVDRSIWSCTL FRATFFNNRLSLLLQSSNAQHSVQVISNSSLGIVG HSPFAYWFDVTNMSTLAILPEYIAFKGDNNRYLN SRFIRTKNRLKFFESDNLLDTGVHHQTHTLRNGNV RFWSIHFGKFWRKSHSEWVFGDSDDKTGGDIN TVFRPIKLDGNKIALLNLGNNGYCMRRVQRGQPR LIDTLIASASSVGPLAVVEVEEAVRQREIYNIHYN MDESRIYNVVPFESGNVTQDTYNLTPLPVIKKVG FKYTQVTRETFDSSHSWKVAYNSKINVTVIPKFV EGEATLDASYEGKVEWSELNETTITQTKDLYIPV PPRTKSVVTLILGRGSVDVPFSYEQRDILPNGRE VHKHMEDGLFSGLNTEVITHTVREISLSTLSESEL EEARRLSEAEEEEAAAPALSSKL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
145	sp_002570_ttmw	0.948838	#N/A	0.953348	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT5G53990.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEVSKKEGSNNLHIGMFPWFAYGHINPFIQLSNK LSSHGIQISFFSIPGNINRIKSSLNLTPPNQLIPLTIP PIEGLSPNFDSSSEVTHETSQLLSLALDQMQPQV KSHLLQLPQIKVILFDYAYHWLPSLASELGIKSV HFNTFPLVVNSYLTVPSRMTDPNKIPTVEEFKKP PHGYPKTSTAASVKTFEAEAYNFLFKCFNGCHYE KLLGFTNRCDAILYRTCNEVEVTSDELDTKWRIW LSKFPEKSVIYCSFGSETSLNNAQIQELTLGLELT GLPFILVLNFGTSNTIDAQNKLEASLPDGFRERIK NRGVLHTGWVQQQNILAHTSVGCFLTHAGFSSV IEGIVNDCHSVN
146	sp_003130_hrrf	0.944787	#N/A	0.9512284	FASCICLIN-like arabinogalactan protein 8 (FLA8); LOCATED IN: anchored to plasma membrane, apoplast, plasma membrane, anchored to membrane, plant-type cell wall; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: FAS1 domain (InterPro:IPR000782); BEST Arabidopsis thaliana protein match is: FASCICLIN-like arabinogalactan-protein 10 (TAIR:AT3G60900.1); Has 14155 Blast hits to 6880 proteins in 856 species: Archae - 79; Bacteria - 4506; Metazoa - 1333; Fungi - 788; Plants - 1999; Viruses - 868; Other Eukaryotes - 4582 (source: NCBI BLink).	MASAKTTTVVFSVLAIFAVSYAHNITEILSSDPSLS QFSSFLTQTKLDDEINSRQTITVLALDNAAMATLT AKHPISVVKNLLSLHILLDYYDPKKLHSIDNGTVLT TTLYQTTGSASSQEGSVNITDLKGGSVGFGSGA PGSKLDSTYTKSVKQIPYNISVLEISAPIIAPALLTA ASSDVNVTGVLEKAGCKTFASLISSSGVLKVYQS QMDKGITIFAPNDEAFKAKGVPDLTKLSNAEIVSL LQFHALSSYSPIGTLKTTKGSLPTLASGVGKYDLS PSAAGDSVTLHTGVDSSRIASTLLDATPVVIYAVD NVLLPVALFGMSPSPAPSPDSDSPSPAPAPEAD SPSPAGTPMSSPPAPSSSESPADSPADSPAGTS ADENTKNGGNERRVPAMFIAVGLSVVSGVISLVL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
147	sp_005150_fhof	0.925743	#N/A	0.973124	cytochrome P450, family 78, subfamily A, polypeptide 6 (CYP78A6); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: oxidation reduction; LOCATED IN: endomembrane system; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: C globular stage, 4 leaf senescence stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, conserved site (InterPro:IPR017972), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome p450 78a9 (TAIR:AT3G61880.2); Has 32637 Blast hits to 32519 proteins in 1695 species: Archae - 48; Bacteria - 3588; Metazoa - 11617; Fungi - 7095; Plants - 9209; Viruses - 3; Other Eukaryotes - 1077 (source: NCBI BLink).	MSTNINIDTIWVFALFSKCNHFSTFNLFISLFVIAIV WLTVALLHWSYPGGPAWGKHRLARMFHWAKK PIPGPRGLPIIGSMSLMTGLAHRKISDMAKTCNAK RLMAFSLGETRVIVTCHHDVAKDILHSSNFADRPI KESAYSLMFNRAIGFAPYGVYWRTLRRVASTHM FSPKQIKASESQRYVIASAMTGYFGTQTGPFQPR DVLRRASLSNMMGSVFGQKCSLVEGPENNNKE SEELREMVGEGYDLLGMLNWGDHLSWLSDFDF QKIRFRCSQLVPKVNRFVGRIITEHRANPVPNSP DFVDVLLNLPSSDKLSDSDMISVLWEMIFRGTDT VAVLIEWILARVVLHPDVQSRVHDELDNVVGRSR AVSESDITSLVYLTAVIKEVLRVHPPGPLLSWARL AIGDTTVDGHHVPAGTTAMVDMWAITRDPEIWA EPDTFNPDRFLGSGAASEISILGSDLRLAPFGSG RRACPGKALGLATVSFWVASLMHEYEWVSSAA DATVDLSEVLKLSCEMAHPLTVEVRRRRASVM
148	sp_008780_frur	0.932882	#N/A	0.920909	uncharacterized protein	MGNSIQGDQTKMSAAKSSLRGCASKFPQPLSSE DLQEAQTLGDTVYQVNHDKLKLQDIIYRSSEGDK DVLRQVFRWDLAPYGDVFQHGFQAHHKNRVHD KTSYDLHSYVQGRGRPLDSRKTSDFAFISTTVSN SWYPPVATGDGNSKTVEVYRYEIYAPGGIWTSL TLGEKNDRKFKRLDEVVFVGGIASQYVRSAQLFR LTGNPDGTTTIERANNVLLINNNFSPQSHPLKLLN IQNPVSDYTDSNRTPLSQTVYQSSNNNNNNTNII NWYANDVANLDVYMSAALRSSHTNEAYYVLQNL VTLLNYDPGGTHDHVVDGPSFVSEMFKSLEYTP FGEHGIDGALKSRDDEAFIFCANLCVLINYAPDTR NDTIVKGPMTVVEMFPFLKGTVFENGIDSACELK SAEDEAYFFKGNLCAHINYGSNPQLITQGQIGQFI PLLKGTVFEGIGVDASYASHAENEVAYYFKYNNY GQINVNVEDGVEYLYGVFVMGDNAPVIHSFVPQ KNRGLDIRSNPGALPSPDEYDYDEL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
149	sp_014050_fdgk	0.9701143	#N/A	0.9351423	carboxyesterase 20 (CXE20); INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: root, flower; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Alpha/beta hydrolase fold-3 (InterPro:IPR013094); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT2G45600.1); Has 9350 Blast hits to 9331 proteins in 1590 species: Archae - 113; Bacteria - 5436; Metazoa - 449; Fungi - 960; Plants - 1345; Viruses - 3; Other Eukaryotes - 1044 (source: NCBI BLink).	MKNPIKLQTFIFFLVTILHNNHLAKAQPPITINPYEF LHIIPNPDGSITRIKEFFPTVPPNTTTSLAISKDVPL NPYKKTWVRLYLPSGLSGPPGYLNNLPIVVFAHG GAFVLLSTATPDIDLFLSNTASKLHVLVVSVEYRL APEHRLPAAYDDVLEALYWVKDKKDEWVGKYG DVSRCIMMGESVGGNIVYNVGLRASVLTRVLKPL VIRGLVLIQPFYGGGNLNRGYKDPNIEVVADLLW NLSLPLGANWNNPYFNPTFGGGSSNLDIIRRLG WRVAFAGCDGDFIFDKQLKVFEFFKGRGLDVIGY FGRGYYHGVFVEDPSESQKIFEFVRYVFSSYI
150	sp_015660_nxer	0.9383109	#N/A	0.9338621	uncharacterized protein	MACEMKAIKFYGELLLFTAINATTSNGADAHIDEF PAQQQHQILPISNILFEAFIPNAARAGGSLTCLGE GEHCWTECGPGLTCKFLGTNSFCGWCPSAGDP CGALDHCCDGYSCDGFFSGTCH
151	sp_017800_nkhm	0.9448032	#N/A	0.9283846	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G19020.1); Has 84 Blast hits to 84 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 84; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	HGSYSNSWADQWDSNNEDQSSSLNNKGNSNN NGVKSGKSMTKKLSKKFGESFNKTKEASANGVE KTKESVSKGMRKAKDTTSGSISWVVQKFRRSSY K
152	sp_022280_xyzq	0.9453525	#N/A	0.933835	Tetratricopeptide repeat (TPR)-like superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide-like helical (InterPro:IPR011990), Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT2G27610.1); Has 41688 Blast hits to 13542 proteins in 246 species: Archae - 0; Bacteria - 10; Metazoa - 149; Fungi - 146; Plants - 40815; Viruses - 0; Other Eukaryotes - 568 (source: NCBI BLink).	MPERNVVSWTALLSGYSQNGFFKDALMVFEDM HHSGVKGNQFTYGSALRGCTKLMCLRSGKQIQG CLQKGRFSSNLFVQSALVDLHSKCGSINDASIIFQ TMPVRDLVCWNTMIDGYAVQGFVNDAFQLFLM MKEGTIPDCFTFGSLLRACTQGNNLSRVFQIHGL MIQHGYGLHDLLNRLLIDAYAKCGSTKHAYTVFK SMMNVFYYIFVELWKTTSCSCTKVFSYFDVALGN ALIDMYAKCGELKEAIRAFDQMEEKNVISWSSLID GYAKHGYGHEAMLLYRKMEHEGLKPNDVTFLSIL FACSHAGLIREGWECFNNMVDKHNIAPTEKHYS CMIDVLARGGQLEEAFNLMQKMNIQPNASHWGA ILGASKVHGDIGLGKLAAKLLFYLDTKNSSTYVAL ASIYAASGLWNDSWGTYKFMEENNLKKSPGCSF DLCEEKRTALLQS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
153	sp_027880_kudg	0.928046	#N/A	0.93443	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: transcription regulator activity, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: sequence-specific DNA binding transcription factors;transcription regulators (TAIR:AT5G50010.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MHNNRKCSNRRGVAPLENQGWVNGQHNVSPA SAFGGSLPSASKMPRPVHDIEFQPSEVCPQNFII VDEDANQSRIMYHPGIAHNLSYPNMDFHETPNG EITQENFGNNFEGGIVSPDMEDSDYIDLLMSLDE NQGESEEEEVSTAGTQGYDGSSSPESFCNYTTK CRKATYPCSVQRASTSGSSHNENKRKEARKMVK VLRGMVPGGDQMNTAAVFDAAVHHLKGLEGKV QRLGLSNHKHFP
154	sp_032700_tnie	0.956098	#N/A	0.944582	Integrase-type DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT5G18450.1); Has 5751 Blast hits to 5635 proteins in 247 species: Archae - 0; Bacteria - 2; Metazoa - 0; Fungi - 0; Plants - 5730; Viruses - 2; Other Eukaryotes - 17 (source: NCBI BLink).	MDNSKKSTLSKPWKKGPARGKGGPQNATCMYR GVRQRTWGKWVAEIREPKKRSRLWLGSFSTAE EAALAYDEAARRLYGPDAYLNLPHLRCSFNPLNK SQRFRWFPSKNFISMFPASSGSSTPSLLNVNVQ PSLHVIHQRLQEFKNKSEPKVVKSDKRVLVEQVS SNEHINWEKPQIDLNEFLEQMGVLKEGETTYTND KVDAVNNHDETAKEIFIEEMNSEWNELVEILRED HQQAESSRSFGDYTMDELITLPPTIWDF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
155	sp_033870_ruye	0.8997733	#N/A	0.8324631	cysteine-rich RLK (RECEPTOR-like protein kinase) 5 (CRK5); FUNCTIONS IN: kinase activity; INVOLVED IN: defense response to bacterium, programmed cell death, response to salicylic acid stimulus; LOCATED IN: endomembrane system; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Protein of unknown function DUF26 (InterPro:IPR002902), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: cysteine-rich RLK (RECEPTOR-like protein kinase) 4 (TAIR:AT3G45860.1); Has 121629 Blast hits to 120099 proteins in 4447 species: Archae - 100; Bacteria - 13629; Metazoa - 44864; Fungi - 10560; Plants - 34243; Viruses - 400; Other Eukaryotes - 17833 (source: NCBI BLink).	KSKADRSHILNISLGVLVAILIVLVALLIYFLCTKRRI TNDSRRENHITNFKNVGESFMSVESLRIDVAILEI ATNNFSEDNKLGQGAFGGYMAPEYIEYGEISTKS DVYSFGVLVLEIISGKKISDRQSELAENLLTLAWM HWSEGLPLELVDTSLGGDSYSSNEIIRCLHLALLC VQGDVEDRPSMLSIVFYLNCSSYLPTPQQPAFLI SRRKGLSGLGSTNDFTDKSTKSGSSSGTSGQHS VKSTSRSNNSFSGIESTMY
156	sp_041880_wyef	0.9482235	#N/A	0.9714423	uncharacterized protein	MARVSFNLVVVPAILAITFMLSTATVLGKVTTTSE EMEMPVGYKMMLEEMALARKVQSNMYDQPAAV RRSQAAMKGLMDVLVSSSSEKAMQCVAAGGYC NLVFGPRCCTNDWACVPPGLVGGVCVA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
157	sp_046690_sjiw	0.936064	#N/A	0.943533	cytochrome P450, family 715, subfamily A, polypeptide 1 (CYP715A1); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: response to salt stress; LOCATED IN: endomembrane system; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 735, subfamily A, polypeptide 1 (TAIR:AT5G38450.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSLVHFVYGFIFISLFLLWKLINACYISPVRTLRKL RSNGLGGPTPSFPLGNLSDMKQMLIKAKSLSSSS LLSSSSESISHDIHSSSLPYFAQWKKLHGKVYVY WMGVEPFVYIADPEFLKEMSGKVFSRNWGKPN VFRSDRVPMLGECGLSMIEGNAWARHRHIITPAF SSTNLKGMVDSMVESTIKMIKNWSTLLGNSSKLE LELDVEREFPVLGGETMAKTNFGIHNSEVENTLL EKLREIQISLFKHTRYVGVPYGKFLSMRQTLKAR KLGKEIDGILLSIINARREIIGVEPRQHEDLLRLLLE GEDHFDGQLTTKELIDECKTFFFGGHETVALSMT WTSMLLAMYPKWQDELREEIKEVIGDKDDIDFTM LARLKKMGWVFNESQRLYPSSPNLQRQARSDIQ VGNVVIPNGTNIWIDVVGMHHDPDLWGNDVNEF KPERFDGDLYGGCKNKMGYLPFGFGGRMCIGK NYALMEYKIVMSLILRRFSFSVAPSYHHFPTYFFS FRPGSGVPLIVKPL
158	sp_053590_purn	0.945215	#N/A	0.972205	AP2/B3-like transcriptional factor family protein; CONTAINS InterPro DOMAIN/s: Transcriptional factor B3 (InterPro:IPR003340); BEST Arabidopsis thaliana protein match is: AP2/B3-like transcriptional factor family protein (TAIR:AT3G11580.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MNHFSLNLPSSDTLWYTPTTMEHHPKSLPFFYH SSLSLSSKHDDVDSDNSSTNNIAATNAIDRSETTS SSPSSADQEYMFEKPLTPSDVGKLNRLVIPKQHA EKYFPLAGSSTVDSSTSSVGDKSTGGVGGGGVL LGFEDESGKCWKFRYSYWNSSQSYVLTKGWSR YVKEKKLDAGDVVLFQRHRLESDRFFIGWRRRG SAPLSQSNPVRYNNNNNNNNNNNNGQNYGVVQ SQSYNGAPQQHQIASFPSYRPDVLHTGTEVRQK GRKSGGGNSKSVRLFGVNLECQIDGEQEEDEEA QTSDGSSQGQGQVFGVGPPAYNYFDYVNNNHY PSSHMDMNYPQDVNQMRM
159	sp_056310_xuqe	0.95598	#N/A	0.938897	late embryogenesis abundant protein, putative / LEA protein, putative; INVOLVED IN: embryo development ending in seed dormancy; EXPRESSED IN: seed; CONTAINS InterPro DOMAIN/s: Late embryogenesis abundant protein, group 4 (InterPro:IPR004238); BEST Arabidopsis thaliana protein match is: embryonic cell protein 63 (TAIR:AT2G36640.1); Has 58153 Blast hits to 33178 proteins in 2866 species: Archae - 714; Bacteria - 20429; Metazoa - 14816; Fungi - 4800; Plants - 4772; Viruses - 420; Other Eukaryotes - 12202 (source: NCBI BLink).	MMERQRFGMKLMSLVIVVAVVVAACNNSVSYAH EVEEEIKESGESWTGWAKEKIQEGLGLKQQGEE LSSGAGEAKEKVTDKASMTKDAAMDKAKEYGRE ASNKASNVKDRVQQQAETAADKAQEAKETVKAD AQEAKEAAARKAQGAKEMAKDEAQLHLETTRNK AQESKETMKANVEETKEAAARVAQRAREMAKDE AKLESESTGNKAQEAKERMTNKVEQTKEATARK AQEAKEMAKDAKDVAEEHSSWVKDKASEGYDA LKKKASDALQTAKEMAQDMTSARYIIDNDDDDDD DEL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
160	sp_064490_jmro	0.9528037	#N/A	0.8338421	Phosphorylase superfamily protein; FUNCTIONS IN: catalytic activity; INVOLVED IN: nucleoside metabolic process; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Nucleoside phosphorylase (InterPro:IPR000845), Nucleoside phosphorylase, family 1 (InterPro:IPR018017); BEST Arabidopsis thaliana protein match is: Phosphorylase superfamily protein (TAIR:AT4G24350.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGVRYELTVVILMVLVHQSVQLRSTHPLHEVVDR INNHGGPYIGLVLAFPAEQVPLLNSGLFVPNSRN PFIQLSGRKFNIGKIKGVDVIYVLTGEQTVNAAITV QSLLQTFNIKGIVHYGIAGCTNTSMSVGDVGVPH SFAFTSSWKWLDFGSEETSEKELKFGAYNHPQT GENLLERIQFSPVQLFTNGKAMEEIFWFPVDAH WYNLATQLQRVKLQQCLNETACLPKEPKVVFGL RGSTANVFMSNLAFAKFLYEEFSISTADEESAAV VHASLSNGVFCVVFRGVSDVAGGEKLALAGLSS LAAINSFKVAAKFVELIGRKKATF
161	sp_066380_dtgp	0.9443681	#N/A	0.9401353	HSP20-like chaperones superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to heat; LOCATED IN: membrane; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20 (InterPro:IPR002068), HSP20-like chaperone (InterPro:IPR008978); BEST Arabidopsis thaliana protein match is: HSP20-like chaperones superfamily protein (TAIR:AT5G20970.1); Has 712 Blast hits to 708 proteins in 96 species: Archae - 0; Bacteria - 37; Metazoa - 7; Fungi - 2; Plants - 649; Viruses - 0; Other Eukaryotes - 17 (source: NCBI BLink).	MMMANTVERKYEDFDPVSDLLSEKECDTLLLYLP GFRKEQLKVQLTTSRILRISGERPTGDNKWKRFN KEVHVTSNVDMKEMTAKFEGGILYIRQPKLITPAE PAATAAPAEPAAPAEQLQDQQGPTGGNMEKKE EGEMDRKEKDEISAESNEERKGKWWMNELREQ VGVRIQLFKELSKPEKLRRLAVYVVLLLILVFQITK MFKSFTKVEEQTVLEFNHQEF
162	sp_068240_kqrd	0.9603211	#N/A	0.9217716	TRICHOME BIREFRINGENCE-LIKE 39 (TBL39); INVOLVED IN: biological_process unknown; LOCATED IN: vacuole; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF231, plant (InterPro:IPR004253); BEST Arabidopsis thaliana protein match is: Plant protein of unknown function (DUF828) (TAIR:AT2G31110.2); Has 1331 Blast hits to 1312 proteins in 27 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 2; Plants - 1329; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink)	MGTTVLNILFLPSLFLLTLQLPQNKAENFCKLANN YTISATNPELAASNGGKCNNFQGKWVFDSKYPL YDSLNCPFIDLEFNCQSRPDKLFQKYRWQPFGC NLPRFNGLYFLRKWRGKKIMFVGDSLSLNQFESL ACMIHSWVPNTKTSYIRGGALRSLTFQEFGVTIM LYRTPFLVDIVREKVGRVLKLDSISGNAWKEMDM LIFNSWHWWTHTGRAQPWDYIQVGNKLYRDWN RWIAFYKGLTTWARWVDLNVNPSKTKVFFVGVS PSHDFGSDWDEPSKSCSSETEPFFATRYPAGTP QEWAVVEKVMSRIRKPVYWLDITTLSQYRKDGH PSKYGGHGKDCSHWCLPGILPDTWNQLLYAALF Q

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
163	sp_074630_yqho	0.994932	#N/A	0.998838	HXXXD-type acyl-transferase family protein; FUNCTIONS IN: transferase activity, transferring acyl groups other than amino-acyl groups, transferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Transferase (InterPro:IPR003480); BEST Arabidopsis thaliana protein match is: HXXXD-type acyl-transferase family protein (TAIR:AT5G07860.1); Has 2450 Blast hits to 2444 proteins in 162 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 125; Plants - 2319; Viruses - 0; Other Eukaryotes - 6 (source: NCBI BLink).	MAKSEQETMAKSEQETSIKLVSECFVKPKYEIKS AKQPYHLGPMDLVMLTIDPIQKGLVFTIKNSPLFL SSESHDNIEIIRTKVVSRLLEKLKHSLSIALVHFYPL AGRFTTQKQPEHNTSLVFIDCNKGPGARFIHATS LDFTISDILSPVDVSIVHSFFDLGEKHVNYDCHTK ALLSIQVTELLDGVFIGFSMSHSVVDGTSFIHFVN TLSEIFKSDDFTTISRAPILNYRPCDIPILKFPFLDV EGFICRAYNPGPLRERIFHFSLNSMLRLKAMANQ ECGTQNVLSSFQALTAVVWRSITRVRNLPKDEQ TTCFMAMGSRTRLNPPLSDDYFGNFMISTKFACK AEELLGNSLGWVAMNLRKIIMSTDEKSILATYKAL ADSPIVIPRETIPGPHGMTRVIIGGSSRFDMYGPE FGLGRALAARMGYGNKDDGKITANPGCEGGGS VDLEIGLRPHIMASLEVDQEFMGFVS
164	sp_084110_jhrq	0.993334	#N/A	0.992848	CD2-binding protein-related; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: suppressor of abi3-5 (TAIR:AT3G54230.2); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MEGPSSSSRKRPFIDEDDDSNKNKQKRVRFPKK KKLKTGEAPEPVAVKDEAEVAGPVDLSDPRTAA KERAKRRGQMTTQLFTEDGEVEVEVAHISAAEV QYQENESFVDDGVQLEPFNLERENEEGYFDEAG NFVEYVPEKEEKDAWLDSLEVYYPKLANKVSTVP NSDDEEQEMSADDTGKMKRRIADVLEPGETVLQ ALKRLKGSTDKRVKMPAEKKVVFDQLTEDAMRL MENGDYNYYHEEREVFVREAEGYEALARARNND VSGIEDDGLDMFADEDNNVAANSEPNNDTSTPS VGTSTQLAVENGSQDSGSTWAGQSDYVYDEAS GYYYSSSLGYYYDPSTALFCSASSGQWYKYNEA TGTYDEVQSETEAPSVAS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
165	sp_084130_zpgz	0.948437	#N/A	0.924015	Leucine-rich repeat receptor-like protein kinase family protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane, membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Tyrosine-protein kinase, active site (InterPro:IPR008266), Leucine-rich repeat (InterPro:IPR001611), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Protein kinase family protein with leucine-rich repeat domain (TAIR:AT1G35710.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSFKALFPLFFCILLQFSFISNFAGAIQQEAEALVE WKQSLHSLNQSFLSSWSLPRDALTAKNSNVITP CHWLGVGCENGSVTRLNLTNVGLTGTIPPEVGY MISLTELRINLNNLTGPIPPSIGNLRGLKTLSVYGN KLSGSLPKEFDNLTNLTLCFLSNNTLSGSLPEKIC QGGILQDFCASNNRFSGTVPKGLKNCTSLTRLRL DRNNFVGNITEDFAAYPVVDYIDLSYNNFEGEISS DWAKCKNMTSLKISDNNITGKIPPELGKATKLHYL DLSSNQLVGEIPKELGNLKSLFNLTLSNNKLSGNI PREVGNLLDLASLDLGANGLNGTIPEEIGNCPKMI YLNLSRNGLNGVIPWQIGSLVSLQVLLDLSRNSL SGEIPLQLGNLDNLEELDLSHNNFSGQIPSTFDQ MQSLISVDVSYNNLEGPLPYGKVFMNSPLVSFME NKALCGNITGLIPCPRALNFSKRKGKNAVALITAPI LGVGVFVCIVVAVALYLLKGGRGVKGEKNRKVG SQRENMFSIWSFDGKLVYEDIKQATKGFDVKYCI GEGGHGTVYKAVLSTGQIVAVKKLKTIQNPSFEN AKNFEAEIEALSKIRHRNIVKLNGFCSHAQHSLLV YEYLERGSLGKLLGIEKEARELDWEKRVIVIKGIA NALCYMHYDCSPPIIHRDLSSNNVLLDRDYEPRV SDFGTARLLSLDSCRNIWLHCSW
166	sp_092350_fytf	0.93054	#N/A	0.922724	Peroxidase superfamily protein; FUNCTIONS IN: peroxidase activity, heme binding; INVOLVED IN: oxidation reduction, response to oxidative stress; LOCATED IN: cell wall, membrane, plant-type cell wall; EXPRESSED IN: root, leaf; CONTAINS InterPro DOMAIN/s: Haem peroxidase (InterPro:IPR010255), Plant peroxidase (InterPro:IPR000823), Peroxidases heam-ligand binding site (InterPro:IPR019793), Peroxidase, active site (InterPro:IPR019794), Haem peroxidase, plant/fungal/bacterial (InterPro:IPR002016); BEST Arabidopsis thaliana protein match is: Peroxidase superfamily protein (TAIR:AT3G03670.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKGSLFSTLILFLGLILMTRVCHCYGGGYGTQLQ QRFYEGKCGGNVDVESVIFGIIKQKIASEPDTVSD LVRLSFHDCFVRGCDGSIFLDGDNTEQKAVINQG IGGIDTINDLKDAVEKVCPGVVSCTDILVIGARAAI SLAGGKYYEVETGRRDGVVSIKSEAQAAIPPPTM PIQQAIALFAQKGLTKDDFVVLLGGHTVGTAHCH SFKDRLYNYRTTQKPDSTISTTLLQNLQKTCPLN DETDNEAFLDQTPDSNFKIDNAYYKQILAQNGVL EIDQNLASSPQTKDMVNTLAQGPDTFQNLFGRA MVKMARIGVLTGTQGQIRKSCGSINT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
167	sp_097540_mrwc	0.9343438	#N/A	0.9347638	Tetratricopeptide repeat (TPR)-like superfamily protein; Has 214 Blast hits to 140 proteins in 56 species: Archae - 0; Bacteria - 4; Metazoa - 151; Fungi - 0; Plants - 55; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	TCSRRKTSMDTLAEGLWGLSDFHEQKGEFAKAI KCLEAICQSPVSFLPIIEVKTRLRIATLLLLRHSHNV NHAKSHLERSQLLLKSIPSCFELKCRAYSLLSQCY HLVGAIPTQKQILNKGLELTKSLGNGFAAKLWYC NFSSQLSTALIIEGDYQNSLATLEQGYKCAAEMC YPELQMFFATSKLHVHLMQWDDVDMVERALAR CDEVWNSFELNKTQQCLGLLFYNELLHIFYRFRI CDYKNIAQHIDKLDAAMKADLQKTEHVQKLTSEL NTINQTLSDPAMKQRDCSMLIEKRALLKQQLVTIT TSNSVAQGYTDSTSVEYERRLSADKLELAPYPID GEWLPKNAVYALVDLTVVMFGRPKGHFKECTKR IQSGTQVIKVELSKLGITDGVREVDLQHSAIWMA GVYLMLLMQFLENKVAVDLTRADFIEAQEALVQM RGWYIRFPTILQACESIIEMLRGQYSHSVGCFDE AAFHFVEAAKLTESKSIQAMCQVYAAVSYCIGDA ESSSQALDLIGPVYRVMDSFIGAREKTSVLFGYG LLLMKQQNLQEARIRLATGLQLTHSNLGNLQLVS QYLTVLGSLALALHDTVQAREILRSALTLAKKLND VPTQIWVLSVLSALYQELGERGNEMENAEYQRK KIDDLHKRFADAYASIHHTALINKDKIEVHQLQES DSKRAMASQSMKVDLDIPESIGILASSTPVKSSSR LVDLDSGRRGKRKI
168	sp_101870_mdns	0.9363362	#N/A	0.9346515	NAD(P)-binding Rossmann-fold superfamily protein; FUNCTIONS IN: oxidoreductase activity, (-)-menthol dehydrogenase activity, (+)-neomenthol dehydrogenase activity; INVOLVED IN: response to karrikin, defense response; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: NAD(P)-binding domain (InterPro:IPR016040), Glucose/ribitol dehydrogenase (InterPro:IPR002347), Short-chain dehydrogenase/reductase SDR (InterPro:IPR002198); BEST Arabidopsis thaliana protein match is: NAD(P)-binding Rossmann-fold superfamily protein (TAIR:AT2G24190.1); Has 78247 Blast hits to 78183 proteins in 3234 species: Archae - 733; Bacteria - 53633; Metazoa - 4843; Fungi - 3816; Plants - 2340; Viruses - 0; Other Eukaryotes - 12882 (source: NCBI BLink).	MAEQSNFLAEKRYAVVTGANKGIGLEICRQLASQ GVIVLITSRDGKKGLEALNDLIKSGISSDNLHYHQL DVTDPMSITALAGFINSKFGKLDILVNNAGIGGFII DYDAIKAIGFRNINYDEMMTQTYELAKECLETNYY GVKRTTEALLPQLELSDSPRIINVSSSTGMLKNIP NERIRGVLGDAENLTEEKVEAILNELLTDFKDGSF KEKEWPSRMAAYTLSKAALNAYARILAKKYPSIIS CVCPGVTKTDMNGNLGQLTVEEGAASPVRVALM PHGSPSGLFYARSEVSSYE
169	sp_114650_ukiw	0.9402233	#N/A	0.9348233	uncharacterized protein	MSFSKFAVFALFMLATSGVIDAKYQPRSSGSIMT EELLEAEVKALASLREVERAVIASKLKQNLMMRT TEDNCIKSGTICSGFGPPEQCCSGACVPHPIIKMF VCQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
170	sp_118730_ygxz	0.921848	#N/A	0.937327	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: basic helix-loop-helix (bHLH) DNA-binding superfamily protein (TAIR:AT2G22750.2); Has 2910 Blast hits to 2902 proteins in 174 species: Archae - 4; Bacteria - 2; Metazoa - 77; Fungi - 51; Plants - 2768; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MEKQALSTMEELLSAALHGEDLCYPLLGKSNTW YPESQWELEYDLVDQFITSPPSYTSNDSSEIISHD DYEFYDNHNNDDGNDNASTVKMLMIENHNNKKV SEKQAVNGGKRKRKRKESQVQDHILAERKRRET LAQLFISLSTLIPGLKKIDKTSILRGAIKHMKHLQER VNVLENLGAQQKNLKSLVVVKKCQLTVDNDDDN NDDNGSFSIIDDNGSSNDAIYLNNSIPEIQVKTTD DTLLLRVHCEQQKDIMTKLFDVVVKHHMSVAHCA VIPFTNLAIDITIVTQMEDGFSMDVKDFVRSIRYILH PAS
171	sp_118740_hmni	0.945663	#N/A	0.95524	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: basic helix-loop-helix (bHLH) DNA-binding superfamily protein (TAIR:AT2G22750.2); Has 2910 Blast hits to 2902 proteins in 174 species: Archae - 4; Bacteria - 2; Metazoa - 77; Fungi - 51; Plants - 2768; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MEKQAFSTMEELLSAALHGEDLCYPLLGKSNTW YPESQWELEYDLVDQFITSPPSYTSNDSSEIISHN DFNDDDNNNNDDGGDNASTVKMLKTTPTGVQE PDRQSENNNKEVTKKQVVNGGGGGKRKRKRQE SQVQDHILAERKRRETLAQLFISLSTLIPGLKKIDK TTILGEAIKHMKHLQEKVNALETLAAQQKTLKSM VVVKKYQLTVANDDNNDDNDNGSRNNAINFNNSI PEIQVKTTDDTLLLKVHCEQQEKIMTKLFDVVVKH HMSVVNCALIPFKNLAQDITIVTKMKEGFDLDVKG FVRSIRSILHPTS
172	sp_123950_hmts	0.968147	#N/A	0.924325	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MTVMARAAVDTSKKGTTFSSFQYSGLKRLGRRN PFMRYGLPLISLTVFGTLGLGHLLQGSKDISKVKD DKEWEIIEERKALSRTGPVNAYNPKNISLEEELKA FQEKVDINSYDYKRIPLPNEGKSAR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
173	sp_128030_ztpg	0.922731	#N/A	0.943825	fructose-bisphosphate aldolase 2 (FBA2); FUNCTIONS IN: fructose-bisphosphate aldolase activity, catalytic activity; INVOLVED IN: response to cadmium ion, pentose-phosphate shunt, response to abscisic acid stimulus; LOCATED IN: in 7 components; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Aldolase-type TIM barrel (InterPro:IPR013785), Fructose-bisphosphate aldolase, class-I (InterPro:IPR000741); BEST Arabidopsis thaliana protein match is: fructose-bisphosphate aldolase 1 (TAIR:AT2G21330.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MATATATATTTSTSLLKSSSAEGLDKSDWIKGQS IRHVAVVKCNPSVNSSLSVKASSYADELVKTAKSI ASPGRGILAMDESNATCGKRLDSIGLENTEANRQ AYRTLLVSAPGLGQYVSGAILFEETLYQSTSDGK KMVDVLVEQNIMPGIKVDKGLVPLAGSNNESWC QGLDGLATRSAAYYQQGARFAKWRTVVSIPNGP SALAVKEAAWGLARYASISQDNGLVPIVEPEILLD GEHGIDRTFEVANKVWAEVFYYLTENNVMFEGIL LKPSMVTPGAECRDRASPQQVAGYTLRLLRQRI PPAVPGIMFLSGGQSEVEATLNLNAMNQAPNPW HVSFSYARALQNTCLKTWGGRPENVQAAQETLL LRAKSNSLAQLGKYTEEGESEEAKQGMFVKGYT Y
174	sp_133220_ytmt	0.952852	#N/A	0.948668	hydroxy methylglutaryl CoA reductase 1 (HMG1); CONTAINS InterPro DOMAIN/s: Hydroxymethylglutaryl-CoA reductase, class I, catalytic (InterPro:IPR004554), Hydroxymethylglutaryl-CoA reductase, class I/II, substrate-binding (InterPro:IPR009029), Hydroxymethylglutaryl-CoA reductase, class I/II, NAD/NADP-binding (InterPro:IPR009023), Hydroxymethylglutaryl-CoA reductase, class I/II, catalytic (InterPro:IPR002202); BEST Arabidopsis thaliana protein match is: 3-hydroxy-3-methylglutaryl-CoA reductase 2 (TAIR:AT2G17370.1); Has 2212 Blast hits to 2210 proteins in 935 species: Archae - 202; Bacteria - 1018; Metazoa - 225; Fungi - 225; Plants - 266; Viruses - 1; Other Eukaryotes - 275 (source: NCBI BLink).	MDVRRRQPPPPPSALKRKTMAGDESPKASDALP LPLYITNGVFLTLFFAVVYFLLQRWREKIRNSTPL HVVSLSEISAIVAFLASFIYLLGFFGIDFVQSFISPR SSYDDDEEVIDDATNEQQILKEDKRISCGIPIPLPE KKLEVHQLPKTLDLTPEDEEVVQSVVSGSTPSYS LESKLGDCFRAAKIRREALQRMTGRSIVGLPLDG FDYDSILGQCCEMPVGFIQIPVGIAGPLLLDGIEFS VPMATTEGCLVASTNRGCKAIYLSGGAASVILKD GMTRAPVVRFGSARRAADLKFFLEDPANFETISL VFNKSSRFARLQEVKCAIAGKNLYVRFTCRTGDA MGMNMVSKGVQNVLEFLQSDFPDMDVIGISGNY CSDKKPAAVNWIEGRGKSVVCEAVIGEEVVKKVL KTDVASLVELNMLKNLTGSAVAGSLGGFNAHAA NIVSAIYIATGQDPAQNVESSQCITMMEAVNGGR DLHVSVSMPSIEVGTVGGGTQLASQSACLNLLG VKGANREAAGGNARLLATVVAGSVLAGELSLMS AIAAGQLVRSHMKYNRSSRDMSKVVASS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
175	sp_133830_rtrk	0.9368835	#N/A	0.9341188	ARA7; FUNCTIONS IN: GTP binding; INVOLVED IN: protein transport, small GTPase mediated signal transduction; LOCATED IN: endosome, plasma membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Ras GTPase (InterPro:IPR001806), Small GTP-binding protein (InterPro:IPR005225), Small GTPase (InterPro:IPR020851), Ras (InterPro:IPR013753), Ras small GTPase, Rab type (InterPro:IPR003579), Rab5-related (InterPro:IPR015599); BEST Arabidopsis thaliana protein match is: RAB homolog 1 (TAIR:AT5G45130.1); Has 26416 Blast hits to 26384 proteins in 719 species: Archae - 27; Bacteria - 133; Metazoa - 13862; Fungi - 3690; Plants - 2969; Viruses - 20; Other Eukaryotes - 5715 (source: NCBI BLink).	MAGSAGKVVKAKLVLLGDMGAGKTNLALRFVRG QFFDCQEPTVGATYFTQTLSVKEATINFDVWDTA GQERYHSMAPLYYRNATAALLVYDISDKHSFVRA KKLVQELLTEETPIPVLSLVGNKADLETEREVATE EAETLSQENGLFFMETSAKTSHNINELFYEIAKRV AEVCPQQPSDIDLQNEIQNRRFSCCSG
176	sp_134000_joqi	0.9373899	#N/A	0.9345632	EMPFINDLICHER IM DUNKELROTEN_LICHT 1 (EID1); FUNCTIONS IN: ubiquitin-protein ligase activity; INVOLVED IN: photoperiodism, flowering, red, far-red light phototransduction, leaf development, regulation of photomorphogenesis; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; Has 121 Blast hits to 91 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 121; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAETTFSKLSDDVTLKIFSLLEDDPRNWARLSCV CRKFNSIITSTCCKSKCLYSFPSISADLLPVSSLDG WASLYKIAVCCPGLQHAGVLFDNSHFGLDRELG PTEEYLHLQTQTRTETETNTTLLSSPPSSSKSPA TTVTEVGEAVVHDDDDDDNNNNNIEFGGWSLYD DLYFDTIYDNSETATPMEVTTGVADLDSKFVVTT NPILEEEDDFESKFVVTKLENDEGVKRRKVNVSR LQRSHLASGVWNLSREQGNKLLASRFRDDCLYI SDWPGCVHIEEKRSYMLFRGVFKNFKRTRVWRT INDGNRSKIVLNCAFCSCNEVWDLHSAFCLKRGF GFHDDGEPVVRAYVCENGHVSGAWTDLPLYT
177	sp_136880_awag	0.9362231	#N/A	0.9328318	Adenine nucleotide alpha hydrolases-like superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to stress; LOCATED IN: vacuole; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: UspA (InterPro:IPR006016), Rossmann-like alpha/beta/alpha sandwich fold (InterPro:IPR014729), Universal stress protein A (InterPro:IPR006015); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G49050.1); Has 2830 Blast hits to 2747 proteins in 619 species: Archae - 112; Bacteria - 1833; Metazoa - 129; Fungi - 32; Plants - 667; Viruses - 0; Other Eukaryotes - 57 (source: NCBI BLink).	MAMSSTEKAVMIVGIDESDESFHALEWTLDKFFT PYAPNFPFNVLLVYAKPTSTTAIGLTGPGAAQVL SFVEADLKRVAARIIIKAKDLCHSKSVHDVSVKVIE GDARSVLCGAVEKNNASMLVVGSHGYGALKRTF LGSVSDYCAHNAHCSVMIVKKPQTDN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
178	sp_138570_dlxk	0.833331	#N/A	0.27341	STARCH EXCESS 1 (SEX1); FUNCTIONS IN: protein binding, alpha-glucan, water dikinase activity; INVOLVED IN: cold acclimation, response to symbiotic fungus, response to trehalose stimulus, circadian rhythm, starch catabolic process; LOCATED IN: mitochondrion, chloroplast stroma, chloroplast, chloroplast envelope; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Pyruvate phosphate dikinase, PEP/pyruvate-binding (InterPro:IPR002192); BEST Arabidopsis thaliana protein match is: phosphoglucan, water dikinase (TAIR:AT4G24450.1); Has 2623 Blast hits to 2596 proteins in 1116 species: Archae - 196; Bacteria - 1989; Metazoa - 24; Fungi - 8; Plants - 221; Viruses - 0; Other Eukaryotes - 185 (source: NCBI BLink).	MSNCLGKNVSLSLLEHQSRINSGTPLKVQATSKLR NSNNTVFSKQFYGKSLRSCTNPKNFMATRRFSY VARAVLAADPASQQADKYNLDGDIELQVSVSPM SSGCISKVEIVVTNSHDDNMVLHWGGIRDSKEK WALPSRRPDGTRVHKDRALRTPFQKSGSGCSLT IEIDDPAVQAIEFLILNEARNIWFKNNGQNFHVRLP RKSEHAQQVPEDLVQVQAYLRWEFRMGKQMYTP EQEKVEYEAARAELMDEVARGVSIQELRERLKK GSNASEDKGQPPSKSESRIPDDLVSIQSYLRWE KAGKPNYSSDEQSREFEEARKELQRELDKGASV DDIRKKIAKGEIKTRVTKQVEKQRNYSSERIQRKK RDLKQVISKHTAEPVDNVSTKVTDPKTLSAVELF SEAKQAQDGDLLRKKMFKLADKQLLVLVTKLDG KTKVCLATDFKEPLTLHWALSAKGNEWSQPPQS ILPPGSSLLDKAAETQFVSVSSDDSSQQFPLQSL EIEIEGDDFVGMPFVLVSSDNWMKDNGSDFVVDI RVKPKQKDAGDGNGTAKSLLEKIASLESEAQKSF MHRFNIGADLLDEAVDAGELGLAGILVWMRFMA TRQLIWNKNYNVKPREISKAQDRLTDNLQNVFRS QPQHQELIRMIMSTVGRGGEGDVGQRIRDEILVI QRNNDCKGGIMEEWHQKLHNNTSPDDVVICQAL IDYIKSDFDISVYWKTLNENGISKERLLSYDRAIHS EPHFRSEQKEGLLRDLGHYMRTLKAVHSGADLE AAISNCMGYKTEGQGFMVGVNIDPLPGLPAGFP DLLQFILVHVEDKNVEPLLEALLEARQELRPLLFK SHSRLKDLIFLDIALDSSVRTVIERGYEELNNAGP EKIMYFISMVVENLALSSDDNEDLVYCIKGWKRA LDMLSSNDKDWALYAKSVLDRTRLALASKAELY QHILQPSADYLGSLLEVDQGALSIFTEEMIRAGSA ASVSALLNRLDPVLRETANLGSWQIISPVEAVGY VVVVDELLSVQNKTYEQPTILVANHVRGEEEIPD GTVAVLTPDMPDVLSHVSVRARNGKVCFATCFD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
179	sp_141040_qmwp	0.955432	#N/A	0.9346776	Integrase-type DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus, chloroplast; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT5G25810.1); Has 5404 Blast hits to 5361 proteins in 232 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 5399; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MEHLPQCSSEFEDRAISPGSPSSGPQSSKSQGS GTRHPVYHGVRKRRWGKWVSEIREPRKKTRIW LGSFPMPEMAARAYDVAAHCLKGCKAQLNFPEE IDFLPRPTSCAPKDIQAAAAKAANAVVEGKVSSS AAADHFDHDHHDDVEVDYDDFWSEIELPELMDT GQGGNFMDCYTSSSLGEFYF
180	sp_151150_kxyd	0.9382239	#N/A	0.9661733	thioredoxin 2 (TRX2); FUNCTIONS IN: oxidoreductase activity, acting on sulfur group of donors, disulfide as acceptor; INVOLVED IN: N-terminal protein myristoylation, sulfate assimilation; LOCATED IN: cytosol, plasma membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), Thioredoxin, core (InterPro:IPR015467), Thioredoxin domain (InterPro:IPR013766), Thioredoxin, conserved site (InterPro:IPR017937), Thioredoxin-like subdomain (InterPro:IPR006662), Thioredoxin-like (InterPro:IPR017936), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: thioredoxin H-type 7 (TAIR:AT1G59730.1); Has 19714 Blast hits to 19583 proteins in 2950 species: Archae - 262; Bacteria - 11021; Metazoa - 1828; Fungi - 880; Plants - 1790; Viruses - 14; Other Eukaryotes - 3919 (source: NCBI BLink).	MGARLSTIPENRGPLVSATAGPGSYPSGYNKPA ASRDRYNWSNSFPAENKAPAAPVCAAFPFPPAN SNPSNNYPAATTTKSSSVAAGCGGDVAGGGRG GAFPVNVKPANTKEVIVFHSSTKWKEYFEASKQS NKLVVIYFTATWCGPCRYMEPSIKEIAAKHTDIDL VKIDVDELFNVSREYGVQAMPTFLFMKKGKQIDK VVGARKEELQRKVEKHKA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
181	sp_151320_gtwi	0.938222	#N/A	0.938231	cysteine-rich RLK (RECEPTOR-like protein kinase) 26 (CRK26); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: stem, root; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Protein of unknown function DUF26 (InterPro:IPR002902), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: cysteine-rich RLK (RECEPTOR-like protein kinase) 29 (TAIR:AT4G21410.1); Has 124218 Blast hits to 122701 proteins in 4579 species: Archae - 106; Bacteria - 13898; Metazoa - 45472; Fungi - 10958; Plants - 35042; Viruses - 435; Other Eukaryotes - 18307 (source: NCBI BLink).	MGKTLFISFASLMVSLWQLIRVATGINFSDEACSV SVANYTINSVYQKNLNLLFSDLSSKTSTQKFYNST IGDDPDKVYGLYQCREDVSLEICRQCIKEATQKIV QVCPLKKEAIVWYLECMLRYSKSSIFSLLDMSVH DYKHSSIGVSNYSQFSPYLASEMDKVIDQAALST SSGHFATRRARWSSYETVYCLAQCTPDLSGRDC NNCLHIALRIMFIANFNNSVAIGIFFPSCQLRYGRV QFYDDLVPKPPAPTSSSSPITMTNP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
183	sp_163870_noif	0.843139	#N/A	0.948336	IQ-domain 33 (iqd33); CONTAINS InterPro DOMAIN/s: IQ calmodulin-binding region (InterPro:IPR000048); BEST Arabidopsis thaliana protein match is: IQ-domain 3 (TAIR:AT3G52290.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGITGELVRSVFSKSRSIKIKTDEINMKFNGVEKK RWKTSVKSYLCGDELFDSVVAEEDSASVRSRSS EATVMQMLQDDGYSTQDIESTKYDNPNRDHNSF SSNSFDEDTAATIIQSAFRGFKVRSLRETTDFQN DFEEGMKSPGRESMATSIEVQTGNSIGVVSIQEE DTSASTKLHHKLKSSAYKPKEDWNDSTLDSTKS KMRIQNRLEAMTRRERALAYAFSQQLRICSKKKH TKSDPDMGWSWLERWMATRQIEVPLVEEEDYY TKQLTENEAVSRTKARPFLRKKFLDLSFEEKESC GSNEIPVQVDTSTNSNGEDKHGYDSPRNRLKNP TSISRTKTVPTYHYQKEYSKVSKKKNRSKETENNT RQKARQSESSSSICNDGASWSPPPDHSVSGHN SKSDP
184	sp_171000_pxii	0.826416	#N/A	0.833033	cytochrome P450, family 96, subfamily A, polypeptide 10 (CYP96A10); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, heme binding; INVOLVED IN: oxidation reduction; LOCATED IN: endomembrane system; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 96, subfamily A, polypeptide 9 (TAIR:AT4G39480.1); Has 28273 Blast hits to 28182 proteins in 1460 species: Archae - 44; Bacteria - 2431; Metazoa - 10638; Fungi - 6078; Plants - 7933; Viruses - 3; Other Eukaryotes - 1146 (source: NCBI BLink).	MAGGDTTSVAFTWFFYLLVKYPQVTAKIREELDA ILITQSNYKDDDFVKNFSDNNKKLVYLHAALCESL RMYPPLLFNHKTSVKSDVLPSGHKVGPNTEIYFD MYAMGRMKSLWGDDCNEFKPERWITKQGNFKL EPSNKLLAFGAGPRICQGRSMTFIQMKIVIAAIIPK YDIVAVEGHQVVPPDVSIVLQMKHGFKVKIVRRPSI PLHDS
185	sp_171320_gmgq	0.856094	#N/A	0.838674	KDSB; FUNCTIONS IN: 3-deoxy-manno-octulosonate cytidylyltransferase activity, nucleotidyltransferase activity; INVOLVED IN: lipopolysaccharide biosynthetic process; LOCATED IN: mitochondrion; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: 3-deoxy-D-manno-octulosonate cytidylyltransferase (InterPro:IPR004528), Acylneuraminate cytidylyltransferase (InterPro:IPR003329); Has 7496 Blast hits to 7495 proteins in 1549 species: Archae - 38; Bacteria - 4007; Metazoa - 8; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 3395 (source: NCBI BLink).	MNPTSICTSSSSSSSSSSTKSWVVHGLVLGAAA AAAIGAHTYFYLRRSRKFRSRVIGIIPARFASSRF QGKPLVHILGKPMIQRTYEQAKLASMLDQVVVAT DDNKIAECCRGFGADVIMTSESCRNGAERCSEA LQKLNKKYDIVVNIQGDEPLIEPDIIDGIVKALQSA PDAVFSTAVTPLKAGDAFDTNRVKCVVDNLGYAI YFSRGLIPYNKSGMVNPDFPYLLHLGIQSYDAEFL KIYPELPSTPLQLEEDLEQLKVLENGYKLKVIKVE HEAHGVDVPEDVEKIEQYMRERNLS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
186	sp_174510_ypax	0.9255632	#N/A	0.9233773	LACHESIS (LIS); CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat, conserved site (InterPro:IPR019775), Pre-mRNA processing factor 4 (PRP4) like (InterPro:IPR014906), WD40 repeat (InterPro:IPR001680), Splicing factor motif (InterPro:IPR003648), G-protein beta WD-40 repeat, region (InterPro:IPR020472), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT2G05720.1); Has 96331 Blast hits to 36444 proteins in 930 species: Archae - 78; Bacteria - 11869; Metazoa - 38307; Fungi - 21059; Plants - 12275; Viruses - 15; Other Eukaryotes - 12728 (source: NCBI BLink).	LLHMIQLLRYETWHSIALSSLFCXILALEGHIKPVY GVDFSPNGYHLATGSEDNTCRIWDLRKKKSLYIL PAHSKLISQVKFEPQEGYYLVTASYDTTVKVWSS RDFKLIRTLSGHEARVTSLDVTGGEIQGCRWFVC GLRVVCVWLEGGFVVVSSGLCLYGDFEWFRDG FMVVGAVYGVGRVVSG
187	sp_177240_yfer	0.937973	#N/A	0.9441167	Mannose-binding lectin superfamily protein; CONTAINS InterPro DOMAIN/s: Mannose-binding lectin (InterPro:IPR001229); BEST Arabidopsis thaliana protein match is: Mannose-binding lectin superfamily protein (TAIR:AT1G73040.1); Has 3242 Blast hits to 933 proteins in 57 species: Archae - 0; Bacteria - 14; Metazoa - 135; Fungi - 2; Plants - 3082; Viruses - 0; Other Eukaryotes - 9 (source: NCBI BLink).	MANQLTNWSFELEDGEKITKVMIRHGFIVDGIGF ETVKNSNSGIPLSSGEFITYSGSYGMYYRETCIT TMKIHTNLRPDGYGPYGRAQGAEGVTDFISPLPL NSSIVGFFGSYGVYLASIGINAERTMITPYGPYGN SESSPNWSIELNEGQRFSKVRISHGYIVDGIGFDI TDQSGKTTPTQLFGGSGGSPSEIDLGVNETITQI SGSYGKYFGKDVAITMIKIHTNKRPDGYGPYGRA QEATDVKPFSTPLPLKGVVFSFQGHYGQQYLNSIG ISDKYKTVEKYGPYGRA
188	sp_178750_skdf	0.944091	#N/A	0.9378948	RNA-dependent RNA polymerase family protein; CONTAINS InterPro DOMAIN/s: RNA-dependent RNA polymerase, eukaryotic-type (InterPro:IPR007855); BEST Arabidopsis thaliana protein match is: RNA-dependent RNA polymerase family protein (TAIR:AT2G19920.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	DQGQLSGPVLVYRNPGAHPGDIHVVNARYVEAL KDFVGNAKYGVFFPSKGLRSMADEIAGGDYDGD MYWVCRNPELLKFFKPSEPWKSNSQPNATKQIS LPSADDVLEMKLFRLYRGVRCQPNYAMGIAAEC WVVYMDRLLTLGDDCADEKATLRRKIDQLIDIYY DALDAPKKGAQVEVPDDLKLKRYPHYLGKGEHR TYRSKSILGEIHDIVDQLSAPIKFWKLECFEADIPG TYKVLWKNHYDKYRPEMQMALNESRNADFVYQ KYRKELYGGVDDLDSSEKAWKDIRLEALTIYHVC YDYAKAAEDPKKCGFAWKVAGDALWKIYLQSTG EKPILFAPSVVRQMMR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
189	sp_179790_koyn	0.8939946	#N/A	0.8339528	wall associated kinase-like 2 (WAKL2); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: integral to membrane; EXPRESSED IN: stem, cotyledon, leaf whorl, leaf; EXPRESSED DURING: LP.06 six leaves visible, LP.04 four leaves visible, LP.10 ten leaves visible, LP.08 eight leaves visible, LP.12 twelve leaves visible; CONTAINS InterPro DOMAIN/s: Wall-associated kinase (InterPro:IPR013695), EGF-like calcium-binding, conserved site (InterPro:IPR018097), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: wall associated kinase-like 4 (TAIR:AT1G16150.1); Has 122559 Blast hits to 120935 proteins in 4599 species: Archae - 125; Bacteria - 14250; Metazoa - 44921; Fungi - 10430; Plants - 34238; Viruses - 502; Other Eukaryotes - 18093 (source: NCBI BLink).	MHLSWLLLLLFSVSAATLAGTDALITPKNMITAAK TSKPGCQQKCGNLTIPYPFGIGQKCYFSGSFGIIC NTSYNPPRTFPVIAWGWDIFDITETQMVIRNSGV ASSCHDSQKNLSSSSTGLLDMWGSPFTVSSTAN KLTVIGCNDYAVFGGSGGDDAAKYQRAGCIAMCN NRAEPVAGSCPGLGCCQTSLPLGWNWQGGFSIT LFSVEDGQLTTIGKSLSDQSCSYAFYAKSGSFIFG GASDLVTNNTDDVRNRIMEQVPMVLDWNVQLKI SCKQVKINSTSNPYLCQSNTKCVDTAEGGYRCS CLPGYEGNPYLQPGCTDINECVSGPNNPCYMTC VNTPGSYKCSCARGYVGDGLKSGTRCIFVNKRA ANLIRITLGLGSSLGVIVLLLVAWWFSTIIKRRKIM KQKVKYFKQNGGLLLQEQMATDDSVVAAIKMFTI DELEKATDHFNQDRILGRGAGALTYLHSSSSIPIF HRDIKSSNILLDDTYRAKLSDFGTSKTVGIDQTHV TTPQVMGTFGYMDPEYHLLGQYTEKSDVYSFGV VLVELLTGQKAVRSALEEEKSLASWFLSHMKNP RLLDIIDSRVLQEDSKEEFLIIAEIAEKCLNQDRNS RPTMKEVLLEIEAVLSLHLPQANEQNWPKSEQIF SGSISDGVVSSSTAFYLENSSLLSSAEVSLLFNSR	MHLSWLLLLLFSVSAATLAGTDALITPKNMITAAK TSKPGCQQKCGNLTIPYPFGIGQKCYFSGSFGIIC NTSYNPPRTFPVIAWGWDIFDITETQMVIRNSGV ASSCHDSQKNLSSSSTGLLDMWGSPFTVSSTAN KLTVIGCNDYAVFGGSGGDDAAKYQRAGCIAMCN NRAEPVAGSCPGLGCCQTSLPLGWNWQGGFSIT LFSVEDGQLTTIGKSLSDQSCSYAFYAKSGSFIFG GASDLVTNNTDDVRNRIMEQVPMVLDWNVQLKI SCKQVKINSTSNPYLCQSNTKCVDTAEGGYRCS CLPGYEGNPYLQPGCTDINECVSGPNNPCYMTC VNTPGSYKCSCARGYVGDGLKSGTRCIFVNKRA ANLIRITLGLGSSLGVIVLLLVAWWFSTIIKRRKIM KQKVKYFKQNGGLLLQEQMATDDSVVAAIKMFTI DELEKATDHFNQDRILGRGAGALTYLHSSSSIPIF HRDIKSSNILLDDTYRAKLSDFGTSKTVGIDQTHV TTPQVMGTFGYMDPEYHLLGQYTEKSDVYSFGV VLVELLTGQKAVRSALEEEKSLASWFLSHMKNP RLLDIIDSRVLQEDSKEEFLIIAEIAEKCLNQDRNS RPTMKEVLLEIEAVLSLHLPQANEQNWPKSEQIF SGSISDGVVSSSTAFYLENSSLLSSAEVSLLFNSR
190	sp_193630_uhio	0.9648012	#N/A	0.9934147	uncharacterized protein	PQTHASTAGYDGSSPHHSYDDDPTQEPIFPPPPT HKKKKTHG
191	sp_197810_urfr	0.9301447	#N/A	0.9238833	polygalacturonase inhibiting protein 2 (PGIP2); FUNCTIONS IN: polygalacturonase inhibitor activity; INVOLVED IN: response to salt stress, signal transduction, defense response; LOCATED IN: cell wall, plant-type cell wall; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: polygalacturonase inhibiting protein 1 (TAIR:AT5G06860.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAQAHLIIVSLLLLVLALFVDLSYSQPITDLCHPHDK KVLLRIKNHLGNPSSLSWDPNTDCAGWTGIQCN SQGHVIRFAINQALDIHGPIPPFLDRLPALTSLYFD GIPNLSGPIPSYIGKLTKLTDFSIIGTNVTGPIPEFL GQLTNLNLLYLYSNKFTGPIPNFLGLLKSLVYLDIS SNLLTGPIPVALAQLTNIEILSLSSNKLSGNIPDFIA QKLKSIKILELNYNSLSGPIPSSLGFLPNISSIGLAG NQLSGPIPKSLGRGNLNFIRLSQNKLTGDASFLFE KANKNVMILEVANNLLKFDFTKVDLAPSLSTMNIG HNMIYGSLPKRFSLFSTDSVDVSYNQLCGPIPNG RRYKRVDPIIFAHNKCLCGGPFPGCK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
192	sp_198340_focw	0.953737	#N/A	0.9402858	HXXXD-type acyl-transferase family protein; FUNCTIONS IN: transferase activity, transferring acyl groups other than amino-acyl groups, transferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Transferase (InterPro:IPR003480); BEST Arabidopsis thaliana protein match is: HXXXD-type acyl-transferase family protein (TAIR:AT1G03495.1); Has 1873 Blast hits to 1858 proteins in 109 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 18; Plants - 1853; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MAPPSSSSTTGSGNGSSFAVNIMASFYISPQQPS TTNSHSIPLTFFDIPWLQYPPLQPLFFFQLPSTPQ SSSSSSSSSSFDHNLYLEFSSTILPRLKHSLASAL QYYFPFSGKLTTTTHTIPNNLVFSTDSSDSVELTV SLCDADFNGLCSFLPRSTHLFQQLVPSLPNIESS NLTTFPAPLLAIQITFFPTSSPGFSIGFASHPVLSD QRTFSNFLYSWASFSKFDNLNISLAPSFPVSDRS VILDPDRLEPLLLEQWLGLESKPTMSTKMKLRPP PAYVRGSLRSTFVMGPSDIANATQWLQTQCEKL NRSYPVLLSPYVVTCAFIWTCFLRARVQNSAVTK AKAKGTMYFGFIAGGITRLPYRVPAKYLGNCVGF GRAAAQREELLKEGEGMLAAADAIGLTIKKLDKD VLGGAEKWIYEWQTLMESEDHIHVVGSPKVNLY ETDFWWGKPKKIEEISTDVTRAISLTQSRDMKRG IEIGLTLPNSIMDDFSSIFTQGLLVFQN
193	sp_199120_jznw	0.9328816	#N/A	0.9231698	Protein kinase superfamily protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; EXPRESSED IN: petal, flower; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT2G18890.1); Has 113420 Blast hits to 112131 proteins in 3664 species: Archae - 114; Bacteria - 13552; Metazoa - 41580; Fungi - 9353; Plants - 32095; Viruses - 456; Other Eukaryotes - 16270 (source: NCBI BLink).	MIPSTAPARIIVATSLEVDDVKDLLSWAINILSHPN DTIIALHVLVGKEPKKLVPKSREYKRFRRAKSFVL SAMGEFAKTCQCKQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
194	sp_199710_jrts	0.948636	#N/A	0.952514	purine permease 1 (PUP1); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF6, transmembrane (InterPro:IPR000620), Protein of unknown function DUF250 (InterPro:IPR004853); BEST Arabidopsis thaliana protein match is: purine permease 3 (TAIR:AT1G28220.1); Has 617 Blast hits to 607 proteins in 89 species: Archae - 4; Bacteria - 97; Metazoa - 10; Fungi - 14; Plants - 377; Viruses - 0; Other Eukaryotes - 115 (source: NCBI BLink).	MYAAHSFITAFAYHRDGTVNRRLVNLADIRSAAS STPIGGVTSFDHTIGIDPLNKLWFRIFFPANSPSS HSLPVIIYFHGGGFVSFSPATYAYDTFCRKVSGD VGAVVISVNYRLAPEHKWPAQYDDGFDVLKFLD DDEMRAKIECFPANADLGRCFLAGDSAGANLAH HVAVRVGGARFKTVRVRGLISIQPFYGGEERTDS ELRTKQWPVLTLSQTDFFWKAFLPNGSDRDHPG SSVFGPKSKDVSGLESFPPTLVVVGGLDLLRDW GMRYYDGLKRSGKKVELVQYSNVGHFFYSAPEL PEYKLFVSKIGEFVENQLRLQGIGISGGPLVMRLY YIHGGKSIWLSSWLETGGWPLMIIPLLVAYFRRL RRRRGGADPTQAAKLIFINTNTFFFAAIVGLVAGA ADYCYAYGVKHIPVSTSSLILATQLAFTAVFAFILV NQKFTPYSINSIALLTFAAIILAVHSNSDKPAGESN GQYYLGFFLTLASSALFALMLPMIELTYKFRVKQSV VVINYSLVMEIQLVLSFSATVMCTVGMLASGDYK EYAREMQDYKHGKAMYFVVLICSAIVWQFFYLGA AGVIFYGSSLLSGVIISVALPITEIAAVFFYHETFQV DKGVSLLLSLWGFISYFYGEAKEIKKQKKRQQRN NTQSIDVES
195	sp_200090_zjji	0.943083	#N/A	0.952732	FUNCTIONS IN: zinc ion binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: intracellular; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), Zinc finger, C2H2-type (InterPro:IPR007087); Has 287 Blast hits to 281 proteins in 115 species: Archae - 0; Bacteria - 0; Metazoa - 116; Fungi - 82; Plants - 70; Viruses - 0; Other Eukaryotes - 19 (source: NCBI BLink).	MAMQVEEGKSGSEFPYWKPVHRKFAPDSPFFS SGDVERQLLDKQVALVLTEDEMQQLLDMEDHER EIFCPIVGCGARLNSLEGFEDHYHSCHTASCSVC HKSYPTQRLLSIHLSEVHDSYFKAKVSKGFSMYE CLVETCGLKFKSYKARHQHLVDKHKFPITYEFYK KAHLSKKQRQKYGRKPALCKEETSNAMPLDNEA DDSMEDLTSAISQLSTSDSVPSSISFGHRHSRGL AFVPRSVQRERRPGYSPSTMN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
196	sp_213110_prhn	0.933338	#N/A	0.866233	vesicle-associated membrane protein 727 (VAMP727); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: vacuole organization, protein targeting to vacuole; LOCATED IN: endosome, SNARE complex, membrane; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Longin (InterPro:IPR010908), Longin-like (InterPro:IPR011012), Synaptobrevin (InterPro:IPR001388); BEST Arabidopsis thaliana protein match is: vesicle-associated membrane protein 726 (TAIR:AT1G04760.1); Has 2397 Blast hits to 2395 proteins in 262 species: Archae - 0; Bacteria - 0; Metazoa - 970; Fungi - 430; Plants - 618; Viruses - 0; Other Eukaryotes - 379 (source: NCBI BLink).	ILNMSPRLHAMTDSNPQLQEMMQNSIVLCQLVFP EMMQVLDCRIRPKIFSFKYAFNCYDLNLVYYKFF WDTSCYFVLVYKQADSFQRQGRQLRRKMWFQN IQLKLTVGGSILFIVIVWLENPQLSDL
197	sp_005860_fgmk	#N/A	#N/A	0.833533	Integrase-type DNA-binding superfamily protein; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT1G15360.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MARPQQRFRGVRQRHWGSWVSEIRHPILKTRI WLGTFETAEDAARAYDEAARLMCGPRARTNFPY NPNAPQTSSSKLLSATLTAKLHRCYMASLQVTKN SAMQAQQRGPGPQFSGTGNVVGPKFGETCAW SPEYGSTDGLVLPNPEPNWTMKVENAQLEPQC HFTMLEDDHIEQMIEELLDYGSIEIAI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
198	sp_008600_oszt	#N/A	#N/A	0.932333	alpha-glucan phosphorylase 2 (PHS2); FUNCTIONS IN: phosphorylase activity, transferase activity, transferring glycosyl groups; INVOLVED IN: response to cadmium ion, response to water deprivation; LOCATED IN: cytosol, chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 35 (InterPro:IPR000811), Glycogen/starch/alpha-glucan phosphorylase (InterPro:IPR011833); BEST Arabidopsis thaliana protein match is: Glycosyl transferase, family 35 (TAIR:AT3G29320.1); Has 5792 Blast hits to 5741 proteins in 1741 species: Archae - 74; Bacteria - 3953; Metazoa - 555; Fungi - 138; Plants - 232; Viruses - 2; Other Eukaryotes - 838 (source: NCBI BLink).	MGTDKANDVTKNGIASKIPAEAHPLSEESTEIASN INYHAQFSPHFSPLKFQPEQAFYATAESVRDRLI CQQWNETYVHFHEANPKQTYYLSMEYLQGRALT NAVGNLNIQDAYADALSKLGHQLEEIVEQEKDAA LGNGGLGRLASCFLDSMATLNLPAWGYGLRYRH GLFKQLISKEGQSEIPEDWLEKFSPWEIVRHDVV FPVRFFGQVQNPDGSRKWVDGEIIQALAYDVPI PGYKTKNTISLRLWEAKASAADFNLFQFNDGQYE SSLQLHSRAQQICAVLYPGDATEGGKLLRLKQQY FLCSASLQDIIFRFKERKDGKTPLKWSEFPSKVA VQLNDTHPTLAIPELMRLLMDDEGLSWDEAWGV TTRAIAYTNHTVLPEALEKWSQSVMWKLLPRHM EIITEIDKRFVELIHSTRPDLEDKVNSMCILDNNPQ KAVVRMANLCVVSAHTVNGVAQLHSDILTSDLFA DYVSIWPTKFQNKTNGITPRRWLRFCNPELSDIT KWLKTDEWVTNLDLLVNLRQFADNEDLQSEWA SAKMASKKRLAQYILEITGVSIDPNSLFDIQVKRIH EYKRQLLNILGAIYMYKKLKEMSPEDRSKTTARTI MIGGKAFATYTNAKRIVKLVNDVGAVVNTDPDVN NYLKVVFVPNYNVSVAEKLIPGSELSQHISTAGM EASGTSNMKFALNGCLIIGTLDGANVEIREEIGEE NFFLFGATADEVPKLRKDRENGLFKPDPRFEEAK DFIRSGAFGSYDYNPLLDSLEGNSGYGRGDYFLV GHDFPSYMDAQAKVDEAYKNQKKWLLKMSILSTA GSGKFSSDRTIAQYAKEIWNIEECPVPSNKE
199	sp_016880_prex	#N/A	#N/A	0.957333	Phosphatidate cytidylyltransferase family protein; FUNCTIONS IN: phosphatidate cytidylyltransferase activity, transferase activity, transferring phosphorus-containing groups; INVOLVED IN: phospholipid biosynthetic process; LOCATED IN: membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Phosphatidate cytidylyltransferase (InterPro:IPR000374); BEST Arabidopsis thaliana protein match is: vitamin E pathway gene 5 (TAIR:AT5G04490.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MLTPPRHESYVNAFSSVKHFSHSLISPSFLTPSIN FSGELRKFRANTSHSLKLRRKKTVAANIMLPDNP VISDICATVITAAIARGVLGIWELTAKYDVFDSKLN RKLVHISIGLVFMLCWPLFSSGYRGAFLACLVPG LNAIRVLLIGLGIYKDEATVKSMSRFGDYRELLKG PFYYASAISFACLVFWRTSPIGIAAICNLCAGDGM ADIVGRRIGKDKLPYNKNKSYAGSIAMATAGFVA SVGYMHYLSAFGFMQENWGLVFRLLVVSLVSSV VESLPISTQLDDNLTVPLASFSTGFLLF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
200	sp_018590_jjld	#N/A	#N/A	0.94828	ammonium transporter 1;1 (AMT1;1); FUNCTIONS IN: ammonium transmembrane transporter activity; INVOLVED IN: ammonium transport, transport, protein polymerization, response to karrikin; LOCATED IN: nucleus, plasma membrane, membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ammonium transporter (InterPro:IPR001905), Ammonium transporter, conserved site (InterPro:IPR018047); BEST Arabidopsis thaliana protein match is: ammonium transporter 1;3 (TAIR:AT3G24300.1); Has 11619 Blast hits to 11599 proteins in 2072 species: Archae - 224; Bacteria - 4688; Metazoa - 462; Fungi - 431; Plants - 505; Viruses - 0; Other Eukaryotes - 5309 (source: NCBI BLink).	MAPSWEESVASSINTIYLLFSAYLVFVMQLGFAM LCAGSVRAKNALNIMLTNVVDAVVGSLSYYLFGF AFAFGTGSEPNPFIGTQFFALKDVPDGKYDYSFF LYQWSFAIAVAGITSGSIAERTQFSAYLIFSFFLSG FVYPVVAHWVWSSHGWLSAGSATLFLGSGAIDF AGSGVVHLVGGIAGLWGSIIEGPRVGRFDAFGKP VPIKGHNATLVVLGTFLLWFGWFGFNPGSFGEIL VPYPTTTDQGNWTAIGRTAVTTTLAGSMSGLVTL FGRRLLIGHWDAMDVCNGVLGGFVAITSGCSVV EPWAAIICGFVAAWVLIGLNMLALKLHFDDPLEAA QLHGGCGAWGLIFTGLFAKEEFIQAYDSGKGDA SRSHGLLLGGGWGLFGSQVLELIVIVGWVSVTM GPLFVVLHKLKLLRISIDDEVAGLDISSHGGYAYH NQDEDERPRRYKDYLQMHN	
201	sp_026640_raed	#N/A	#N/A	0.94828	LOCATED IN: plasma membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF3754 (InterPro:IPR022227); BEST Arabidopsis thaliana protein match is: aminopeptidases (TAIR:AT5G13940.1); Has 308 Blast hits to 303 proteins in 113 species: Archae - 0; Bacteria - 164; Metazoa - 10; Fungi - 0; Plants - 111; Viruses - 0; Other Eukaryotes - 23 (source: NCBI BLink).	MGKKKKEFIRLDRESVIPISKPRLIHGLTQLFESM SEREEFSRFCKRVEYTIRAWYLVKFEEMMQLFNL FEPVAGAHKLKERNLSEDEIDELEQKFLVLLFQM MNKSNFKIVTDQEFEVAASGSYCLNLQIKVDEDK LDSTLLRSYFSKHPQDHLPYFADQYMIFERRGCGL DQMTGYFIQWKVDALISRTVQGFLRLTGLKKIFS RKGRLWYVRNQELLEGIVTECGEEDEYDDNDGDE HSLLIERIRVQNMDLSICNLFSKTTIQEPTFEKIIIVY RLHSTKDGRAIYVKHFKHIPMADMEIVLPEKKSPG LTPSDWVKLLVSAAIGLVTVITQLCQTKANIKVLA TVCSAVVAYCAKIYFTFQANLKDYQSLIVQSMYD KQLDSGRGTLLHLCDDVIQQEVKEVIVAFFVLIKY GKLSRQDLDQKCEELLSDQFNEKCNFDIDDAIHK LEKLSLVYKDEMEQYSSVDVGKANEFIRITTEEIV TIACQGDQ	

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
202	sp_027050_wigk	#N/A	#N/A	0.9338471	myo-inositol monophosphatase like 2 (IMPL2); FUNCTIONS IN: histidinol-phosphatase activity, 3'(2'),5-bisphosphate nucleotidase activity, L-galactose-1-phosphate phosphatase activity, inositol or phosphatidylinositol phosphatase activity, inositol-1(or 4)-monophosphatase activity; INVOLVED IN: sulfur metabolic process, histidine biosynthetic process; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histidinol-phosphate phosphatase, putative, inositol monophosphatase (InterPro:IPR011809), Inositol monophosphatase (InterPro:IPR000760), Inositol monophosphatase, metal-binding site (InterPro:IPR020583); BEST Arabidopsis thaliana protein match is: Inositol monophosphatase family protein (TAIR:AT3G02870.1); Has 15134 Blast hits to 15130 proteins in 2267 species: Archae - 157; Bacteria - 8374; Metazoa - 557; Fungi - 304; Plants - 238; Viruses - 0; Other Eukaryotes - 5504 (source: NCBI BLink).	MYSQTHLLSQSPSFLSLSSNSSAFIPNHREIQSLH FSSTYSLSSSRYSLSSRTNLIPTPMAASNPDLLNH QTDGDLLGFDNGNVDRFSDVANELADASGEVIR SFFRKKFEILDKDDLSPVTIADRTAEESMVSIIMKH FPSHAVFGEENGWRCKEKHSDYVWVLDPIDGTK SFITGKPVFGTLIALLYKGKPILGIIDQPILRERWIGI NGKRTTLNGQEVSTRSCTDLSQAYMYTTSPHLF AGVAEDAFIRVRNKVKVPLYGCDCYAYALLASGF VDLVVESGLKPYDFLALIPVIKGAGGTITDWKGEE FHWEASPDSQATDFNVVAAGDPNLHQQTVLSLE WK
203	sp_031320_wiin	#N/A	#N/A	0.8833132	GDSL-motif lipase 5 (GLIP5); FUNCTIONS IN: lipase activity, carboxylesterase activity; INVOLVED IN: lipid metabolic process; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Lipase, GDSL (InterPro:IPR001087); BEST Arabidopsis thaliana protein match is: GDSL lipase 1 (TAIR:AT5G40990.1); Has 3342 Blast hits to 3298 proteins in 168 species: Archae - 0; Bacteria - 231; Metazoa - 0; Fungi - 23; Plants - 3083; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MLLESLISLIYFLLISFPSNINCLISSNHHKHHPPTS LFIFGDSILDVGNNNYINTSTLDQANFWPYGVTFF HHPTGRFSDGRVISDFIAERANLPSIPPYLQPGK WRFYDGVNFASAGGGALVGTFKGYVIDLHMQVK HYKTVEKWLKNKLGNVKAKTKLERSVYLFSIGTN DYMSLFLTNSSTHLTHNKSQYVSMVINNITSVVN EIYKTGGRKFGFLNLFPVGCFPALRMLNPQGND ECQTEVSSYIKLHNKAISIALKDLAKDLPGFKFALF DYYTSTLQRVNHPSKFGYKEGKTACCGTGRFRG VSSCGGRRPVKEYELCKNINDHIFWDSLHYTEKA SKQITDEMWSNINYIHGSYTIKDLFHLPS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
204	sp_032210_afjg	#N/A	#N/A	0.922365	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity, iron ion binding; INVOLVED IN: oxidation reduction; LOCATED IN: cellular_component unknown; EXPRESSED IN: 7 plant structures; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Isopenicillin N synthase (InterPro:IPR002283), Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT3G47190.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSESNIIENHSYPPIFRSPPNDTTRKLDLKDEQGE GAESTGTLDPLPLPLIDFGCLDDEKLDTACKE WGIFRLENHGVPSSLLRQLQEHAKTLFRLPFEAK QGLAKRCPIKYFWGTPALTSSGVALLRTSDALNM DWVEGLNVPVSQLHQHLQHIHHDPLINDFRDLLE DYRGHLTRIASTLFEAMINNLKLDLQQVKPCLDES TGLIRVYRYPRHTINEHVKGLHEHTDSSIISILSED QVGGLQLLKDDKWLKVEPIPNTLVVNLGDMMQV LSDDEYKSVKHKVELNKNKERVSIGYFVFPEENT LINSSKYKPFSYNQFRAQVQQDIKLHGTKIGLQNF KLPHTSQTV
205	sp_033940_zwyp	#N/A	#N/A	0.9344137		

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
206	sp_034930_tffr	#N/A	#N/A	0.923881	SUPPRESSOR OF SALICYLIC ACID INSENSITIVITY 1 (SSI1); FUNCTIONS IN: transferase activity, transferring glycosyl groups, starch synthase activity; INVOLVED IN: amylopectin biosynthetic process; LOCATED IN: chloroplast; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Glycogen/starch synthases, ADP-glucose type (InterPro:IPR011835), Starch synthase, catalytic domain (InterPro:IPR013534), Glycosyl transferase, group 1 (InterPro:IPR001296); BEST Arabidopsis thaliana protein match is: starch synthase 2 (TAIR:AT3G01180.1).	MESLSLQSFHVKGLNYRSKFNLANNPSEFRAINQ VGFDHFLRQRCKRLCFSVRCQSFSGKNEGGVSS SADGSSSAIEEQRNEGHFGQLIGPVKDASGSIVS FNLYSQSGDGKSNVFLDLEKPNENDKEEKISSGS PSSFVFERGVIKESEVEKILEETIVEGEDKTVAKS KLSVVFVTAEAAPYSKTGGLGDVCGSLPIALAER GHRVMVIAPRYIHGTAADKVYAGAFDANCRIKVN CFGGEQEVAFFHQYRDGVDWVFVDHPSYHRPG NPYGDSFGAFGDNQFRFTLLCHAACEAPLVLPL GGYTYGEKCLFLVNDWHAGLVPVLLAANYRPYG VYKDARSILVIHNLSHQGVEPAVTYDNLGLPPQW YGPLEWVFPEWARAHELDKGEAVNILKGAVVTA DRILTVSQGYSWEITTVEGGYGLHELLTSRKFVIN GIINGINVSEWDPSNDAHIAAPYSIDDLSGKAECK AALQKELGLPVKPDDCPLIGFIGRLDYQKGIDIIQAA MPELLGDDVQFVMLGTGVAKYESWMRETEARY KDKFRGWVGFSVPISHRITAGCDILLMPSRFEPC GLNQLYAMRYGTIPVVHSTGGLRDTVETYNPYAV GPDGSGTGTGWGFAPLAKDSMLSALRNGLKTY REYKGSWLGLMKRGMQQDFTWDSAAAHYEQV FEWAFTDPPYC
207	sp_035120_edcp	#N/A	#N/A	0.95438	PR5-like receptor kinase (PR5K); FUNCTIONS IN: transmembrane receptor protein serine/threonine kinase activity, kinase activity; INVOLVED IN: response to fungus; LOCATED IN: endomembrane system; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Thaumatin, conserved site (InterPro:IPR017949), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Thaumatin, pathogenesis-related (InterPro:IPR001938); BEST Arabidopsis thaliana protein match is: receptor serine/threonine kinase, putative (TAIR:AT1G70250.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MILIAFIIIIYAKQRRRKRKFSLLWTKGTMMNQINVE AFLKSYGSLAPKRYTYAEIRKMTNNFEDKLGEGG FGAVYKGKLPDNSRFVAVKKLNKLKGNGDDLNN EVVSMGRTNHVNVVTLLGFCFEGNKRALIFEFFP NGSLEKFIVTAETDHSLTSETLFSIAVGIAKGLDYL HRGCNTRILHFDIKPHNILLDEELRPKISDFGLARL CPPNRKESLISMSGARGTIGYIAPEVFCRNFGGV SHKSDVYSYGMMLLDMVCGRNNIVSEAKGSSEV YFPEWIYSQLEIEDELENDIVVHGMIKDDEEKEVR RN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
208	sp_037530_uqpo	#N/A	#N/A	0.548847	DEFECTIVELY ORGANIZED TRIBUTARIES 3 (DOT3); FUNCTIONS IN: signal transducer activity; INVOLVED IN: in 6 processes; LOCATED IN: cellular_component unknown; EXPRESSED IN: leaf primordium; CONTAINS InterPro DOMAIN/s: NPH3 (InterPro:IPR004249), BTB/POZ fold (InterPro:IPR011333), BTB/POZ-like (InterPro:IPR000210); BEST Arabidopsis thaliana protein match is: Phototropic-responsive NPH3 family protein (TAIR:AT5G64330.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKKPVPAAKAEIQENNSNDAADQITLPTEINNVD YTFQKKDHSWFVTSDIPTDLLIQIEENTFEHVHKYA LLTKSGYIGRTEFQPTSSGLENNIKLDNFPGGPVT FKEVLKFCYGIPVDLSANNVAPLRCASEFLEMTE EFEDGNLISKSEAFLTFIVLSSWRDSVIVLKACEN LSPLAENLQIVRRCCDSISRKAFQDKNELIDDKA WWFDDVATLRIDHFGRIIAAMKPKGIKPDILGSCI KHYAEKWLPSMDGEVEGQRRHWYGKNELQISII CGANHEDDIGQNKDQKIIIESLISILPPQKEAVSCK FLLWMLKRAMICYATPALISELEKRVGSVLEDAN VQDLLIPRCTNKDRGGLNSPEERTMHSIDTVQRII EYFLMHEHQKKTEDLDVSKLLDNYLAEIATDPNL SISKFQVLAEALPANTRVCHDGLYRAIDTYLKTHP SLPEQERRRLCRIMNCQRLSLDACMHVAQNDRL PLKIVIQVLFSEQVKMRTAIQSKEQTIEGNTREND ESWSSAKKEIMNLKTELEKVKERMSELQRDYSE LQREYIMQINKPRRNSSWIFRWKKKMKSISLFTEK AEEEIGEDQETPNQEGSTPNIIRRLSIT
209	sp_038280_etzk	#N/A	#N/A	0.538318	Protein kinase superfamily protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation, cell wall macromolecule catabolic process; LOCATED IN: plasma membrane; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidoglycan-binding lysin domain (InterPro:IPR018392), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: protein kinase family protein / peptidoglycan-binding LysM domain-containing protein (TAIR:AT2G23770.1); Has 102109 Blast hits to 101070 proteins in 4036 species: Archae - 73; Bacteria - 10993; Metazoa - 38718; Fungi - 7726; Plants - 30221; Viruses - 349; Other Eukaryotes - 14029 (source: NCBI BLink).	MFPKISTFKSFQISHCLILLLFLNTIHHSNSQPPTN GYTCTTTTNTSNNIYPCQTYVLYRPKPPEYLDLA SIADLFNVSRLSISKPSNISNPSFSLLPDQPLFIPIT CGCNPVRNQTLKSLSFARMTYTFKPDDTFYFVST HVYGNLTTYESVEVVNPELDVLNLQIGSDAYFPIF CKCPNKTQLGNNVTCLVSYVFQPDDTLDSIAELF KVTKQSIADVNANNLGAFDTVFVPVSKLVTLHQP PGVNQTNAGGSNSTNNQTTAMNGNSDDHKKAV VGLGVGLGVSVLLLVLVSGLLIYREIEYRKKLKER ELEVISRPENGGMGDQFRIIKQLEQNLLLADVSD TLDKYKIYEIDELRKATGGFGHKFLIQGSVYKGCI DGKDFAIKKMKWNAYEELKILQKVNHGNLVRLEG FSIDPEDASCYLVYEYIENGSLNSWLHENKRKLN WKARLRISIDVANGLQYIHEHTRPQVVHKDIKSSN ILLDSNMRAKIANFGLARSGCNAITMHIVGTQGYI APEYLTDGVVSPKMDVFSFGVILLELLSGREVIDE EGKVLWVSAHEIWECSDEGEKARKLMEWIDKFL QKESCSVDALTNIMTVAVACVNKDPSKRPSMVD VVYALCKSDDLFCDISGNEFSSSPITAR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
210	sp_049620_ccgu	#N/A	#N/A	0.833742	villin-like 1 (VLN1); CONTAINS InterPro DOMAIN/s: Gelsolin (InterPro:IPR007122), Villin headpiece (InterPro:IPR003128), Gelsolin domain (InterPro:IPR007123); BEST Arabidopsis thaliana protein match is: villin 4 (TAIR:AT4G30160.1); Has 3095 Blast hits to 1746 proteins in 186 species: Archae - 0; Bacteria - 6; Metazoa - 2441; Fungi - 102; Plants - 219; Viruses - 0; Other Eukaryotes - 327 (source: NCBI BLink).	MLEHDVHYWLGKDANEVESLMVSDKALELDLAL GSCTMQYREKQGQESAKFLSYFKPCVIPCEGVY TSGLERSNRGTYHITLLA
211	sp_054040_cmic	#N/A	#N/A	0.833718	PAP/OAS1 substrate-binding domain superfamily; FUNCTIONS IN: nucleotidyltransferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Nucleotidyl transferase domain (InterPro:IPR002934), PAP/25A-associated (InterPro:IPR002058); BEST Arabidopsis thaliana protein match is: Nucleotidyltransferase family protein (TAIR:AT2G40520.4); Has 391 Blast hits to 365 proteins in 105 species: Archae - 0; Bacteria - 0; Metazoa - 57; Fungi - 102; Plants - 177; Viruses - 0; Other Eukaryotes - 55 (source: NCBI BLink).	MDTQDLRSPLPSSFSSSSSSNSDVIPFTIGATRW QVGDRTAQQILAVIQPNAESENRRKEIIEYLQRLIE DALGIKVFPFGSVPLKTYLSHGDILTAVSHFSTS QDLAADICCLLENEKKNNANVQDLQVKDVICVPA QVRVVKCTVGGICVDISFNQTAGLCAFYFLEQVD RFIGKDHLFKRSVILIKAWCYYESRLLGGFHGLIS TYALETMVLHVINLFHSSLPCPLAVLYKFLEYYST FDWNSYCVDINGLVIIPFLPEIVVQSPQDCDDVLL TDTFLRSSREAILSSMGVGRTNDSAFSVKFLNVL DPLKEDNNLGRSVCKGNFFRIKSALTYGAKQLGG VLMLPLQNIQNGLEKFFKDTLERNGKGRRADVIV PVDEYGSNTSTGNSGHDHLLSGIHYGLWFHDYG MMYPVLPETSENSRVWSSQWVVDMGQNVQLDH SMYNNVNQSPPSSVNRVTSFPQTNAASVKEKW KSRGTGTYIPNLEHLMRRNIVLEARKNEHECEKE QGKGKGKEVESVDSPACTPREANLSSDEGANTD VFNFSIEDFPVLPGSKPITIKVHQLEAAPEPEQVKE MSVSLDSLEFDNKHSSSRSQSLTEAGKQSDSDV SAILDLVVDPTTVAVLEDQESSKSEEKMVKPLEL ADDNEFPPLSRPPEPVITLNKKEFPPLRASLKSQ KKGKWKKDFK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
212	sp_055020_skro	#N/A	#N/A	0.957413	Concanavalin A-like lectin protein kinase family protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Legume lectin, beta chain (InterPro:IPR001220), Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Concanavalin A-like lectin/glucanase, subgroup (InterPro:IPR013320), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Concanavalin A-like lectin/glucanase (InterPro:IPR008985); BEST Arabidopsis thaliana protein match is: Concanavalin A-like lectin protein kinase family protein (TAIR:AT3G45330.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MFQVMAETCSYLHKKIALLFFFVALTFAQDEKQF MYNGFQGAKLHLDGLANIHPNGLLQMTNSSLQG SGHVFYPQPLKFNTSRSFSTTFVFAINTDVPGHG GHGIAFVISSSMNFDHAVPAEYLGIFNISNNGNNS NHVIAVELDTVQNAVFKDIDGNHVGIDINGLESVG SASVAYYSDEERKNMSLQLSGGTPMQIWIDYDG AKMLMNVTVAPLKHPKPKKYLLSTPVNLSPALLD PVYVGFSSSTGLTATEHYISGWSWSQIGQAQDL DLTKLPSFPKIRSQRKKLKVLFIVLIVVLSLLLEMIA GVAFAVWRKKYEELKEPWEQEYISHRFSFKDLY TATKGFKDSELLGFGGFGKVYKELSITGKPTTAT DVFSFGIFLLEVACGRRPNSLQEFSREDFVLVDW VYDCWKKGVILETSDPTLEGNHQVEEMELVLKL GLLCSHPKPEGRPSMRNVVQFLHGDASLPDIPS DYDHQSNSYFGGSWRISSSLPSLSSLGNISRQG CDLSEALFFRFSNIL
213	sp_057480_jiom	#N/A	#N/A	0.934132	Glycosyl hydrolase superfamily protein; FUNCTIONS IN: cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: defense response to fungus, incompatible interaction, response to salt stress; LOCATED IN: cell wall, plasma membrane; EXPRESSED IN: 11 plant structures; EXPRESSED DURING: LP.06 six leaves visible, LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage, LP.08 eight leaves visible; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, family 17 (InterPro:IPR000490), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: beta-1,3-glucanase 1 (TAIR:AT3G57270.1); Has 2169 Blast hits to 2154 proteins in 141 species: Archae - 0; Bacteria - 6; Metazoa - 5; Fungi - 22; Plants - 2121; Viruses - 0; Other Eukaryotes - 15 (source: NCBI BLink).	AQVGVCYGMLGDNLPTEKDVVHLYKSNGIRAMR LYSPNLATLQALQGSGIKLMLDVPNETVPSLASS DPSSAMQWVTTYVVPFASSIKYIVVGNEIHPSDK QASSVLPAMQNVLNALKQNKLGDLIKVSTSIDTTL IVNSYPPSDGQFNDTSSSYTAPIIKFLSTNKSPLLV NVYTYFAYVSNQRDISLNYALFTSPGTVVTDPNN NKNYKNLFDAMVDSVYAALAKAGAPNTAVVVSE TGWPSDGGDAATVNNAGTYYRNLINHVEQGTPL RPGQAIETYLFAMFDEDKKIGANTERNFGLFTPT KHPKYGQLNFS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
214	sp_063280_hapr	#N/A	#N/A	0.9448444	uncharacterized protein	MASEQPASTSGRCPGKEPVTDDIDMATGEYSDG DSIFEDSAGKDVDEHRGGDGDEAGVSGRNDTD VGNDGIGSDDDDPADVIVAEVLADEAKEQPYFTD DSTDVHRREELPPLDSLPRTKTLDRDIFGFDENW RPVCTQPPPPKGKYDTISTYSTRASSRRSAVTVA KNRVGTGSKTKSSSASISKTAQPSNMQSNPTGR GGKSRKRSSLTSGRSDPNLKSKAAGGDDVSAAP SIEILDIEGGGGENEPPLQHCATSAAVTGGNSSM PPVSRSENRSGDSRLNYHLPLRTGGWIEDTPFKI PPAMKSWFGSPGGEPTDQFYPNIDLLCGESVVQ SSIDLYYVYRWNQLQLAQNSIDIAELKKRAEAAEA ALKDSAKKLETASSDLEAVNKRLEDVEPKLKAET ERADDLTQQLIDLNRGLPGVKKTAAKRAVDKLLQ SD
215	sp_071600_ifwj	#N/A	#N/A	0.8335858	uncharacterized protein	MENPHFLSLTLVDLSLSSTAPLSLISANADHPLRR RSDSCCCGVAALLLLHRVSPLSPSPSILLRLTPPP WWSLLPPLLLLPHKWVTSLS
216	sp_075610_odsj	#N/A	#N/A	0.9328285	cytochrome P450, family 707, subfamily A, polypeptide 2 (CYP707A2); FUNCTIONS IN: oxygen binding, (+)-abscisic acid 8'-hydroxylase activity; INVOLVED IN: in 7 processes; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 707, subfamily A, polypeptide 3 (TAIR:AT5G45340.1).	MQYSPLLLVVSPTNNPFLISLVLFLFSLVILTLSQL LCYFYPKHKRLQLPPGSMGWPYLGETLKLYTEN PNSFFSNRQKRYGDIFKTHILGCPCVMISSPEAAK IVLVTKAHLFKPTYPPSKEKMIGPEALFFHQGDYH SKLKKLVQSSFLPCTLKASVSDIENIVLNFLPTWE NNTINTLHEMKKYAFDVAIISAFGDNKELEIEGIKN LYHSLEKGYNSMPLNFPGTPFHKSMKARKLLNET LRKLIARRRKNNNIDGGLLGVLLGSREQKTGKNN ELSDSQIADNVIGVIFAAHDTTASVLTWLLKYLHD NHHLLQAVTKEQGGIKQRLIEANRGLTWEDTRN MPLTSRVIQETLRTASILSFTFREAVHDVEFEGYLI PKGWKVLPLFRSIHHCAEFFPHPEKFDPSRFEVP PRPNTYMPFGNGVHSCPGSELAKLEMLIFLHHLT TTYRWEVVGDEDGIQYGPFPVPKRGFPIMVHPQ K

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
217	sp_076640_oyhh	#N/A	#N/A	0.3288398	Protein kinase superfamily protein; FUNCTIONS IN: protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 7 plant structures; EXPRESSED DURING: C globular stage, F mature embryo stage, petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT5G61570.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKQTQFLKFLFLLIPFTFIIFILIYLLKRKNPEKNQP SETGDPELGNTINKNNKEEEEEEEEELMCFQGG EDLTVPDILDAPGEVIGKSSYGTLYKANLYRNNN NNNSILLLRFLRPACSGTIREVIPVVHAIGLVRHSN LVPLKAFYMGPRGEKLLVHPFYAHPNLAQFIKDS NNESHKWDVIYSIALGLAKGLAYLHTGLHKPIIHG NLKSKNVLLGSDFQPYISDYGLHLVLNPTAGQEM LETSAAQGYKAPELMKMKDVSESTDIYSLGILLLE LLTGKEPVITANSSDSEQLYLPNVIRRAILDHRIND MYHPQIVSNQTGKQRHKTEDCALKLVQLAMSCC SASPSLRPEASEVARKVEEIGKEQALQHKRT
218	sp_077030_crpg	#N/A	#N/A	0.3340436	Tetratricopeptide repeat (TPR)-like superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885), Tetratricopeptide-like helical (InterPro:IPR011990); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT1G73710.1); Has 56373 Blast hits to 14756 proteins in 361 species: Archae - 8; Bacteria - 288; Metazoa - 580; Fungi - 796; Plants - 52865; Viruses - 0; Other Eukaryotes - 1836 (source: NCBI BLink).	MFSKAQCFDINCFFKSPCHLHTSFNIKFSVSPIET HDSSKKPRDSNVLEIPNKGYQKFRRFEKQNANS CGVIDKQSSLSKNPHGSVDELGRLQSGNSRNQQ KKINGEVEFAGGSGNFPAKKVHTRCSTKWASYG GCIPSILLALDSISDLDEAFRPWEERLSNKERSIIL KEQSSWERAMEIFEWFKKKGCYELNVIHYNIMLR ILGKARKWVELESLWDEMKMKRIAPINSTYGTLID VYSKGGLREEALVWLQRMNEQGMEPDEVTMGI VVQMYKKEGEFRKAEDFFKMWSSDKSKKEKNT FQGHKMGAMKVKSGDQLSLSLYTYNTLIDTYGK SGQLQEMSQTFQKMIDDGVIPDTITFNTMIHIYGN IGRLEEVNKLVRKMAELNRSPDTRTYNILISLYAK RDDINLAFSYFARMKQDYLQPDPVSYRTLLYAFSI RQMVFEAEMLISEMDKRGLEIDEFTQSALTRMYI DAGMLEKSWSWFERFHLEGNMTTECYSANIDAY GERGYVKEAEEVFVCCQEKNKIGVLEFNVMIKAY GIGKSYDKACQLFDSMGKLGVFPDRCSYISLIQSL AAADLPHIAKPYLLKMQEAGLVNDCVPYCALISSF AKLGQLELAKSLYKEMIGYNVRPDVVLYGVLINAS AESGNVSEAMEYVDALKKADLSLNDVICNSLIKLY TKVGYLEEAVETYRLLRSFEKGADVYSSNCIIDLY SKRSMVTEAEEVFDDVRERGEANEFSFAMMLC MYKRVGRLEEAMEMAEKMQELGLLTELLSYNSV LALFAMVGRLKEAVSIFKEMISLEIEPDDSTYKSL GFVLVKSGVPKQAVAKLEVARKKDTKSGLQAWV SALTSVVCRDDYGNYDSD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
219	sp_079960_puqw	#N/A	#N/A	0.937341	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: basic helix-loop-helix (bHLH) DNA-binding superfamily protein (TAIR:AT2G22750.2); Has 2910 Blast hits to 2902 proteins in 174 species: Archae - 4; Bacteria - 2; Metazoa - 77; Fungi - 51; Plants - 2768; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MEIQTVPSLINNSHNEYNQSETTEAASFDEQLTA ALLGEDFPYPLPAKNCTTYPDQSHWDFEYDLVD QFITPPSYSSNSSASYDYDNNNNYDDVNDNTM VSPGVQEPGRQSKKRKEVITEKQAVNGGTRKRK RHPSQVQDHILAERKRREVLGHMFISLSTMIPGL KKIDKTSILGEAIKHMKHLQEKVNVLEKVSAQQKS ARSSMVVVKKYQLGVENDDSDDTGNGSSSIVSD NSASDSGGGGGGCGSNGDSSNNAMNLNNSIPEIE VKITDETLLLKVYCETQNGVMAKLLKVVENHHLS VTNCSLLPFDKLALDITIVAQIQEGFDMNVKDFVR SIRSVLRPAGI
220	sp_080330_ahoq	#N/A	#N/A	0.947677	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: pentatricopeptide (PPR) repeat-containing protein (TAIR:AT2G22070.1); Has 43848 Blast hits to 14238 proteins in 253 species: Archae - 0; Bacteria - 12; Metazoa - 20; Fungi - 118; Plants - 43065; Viruses - 0; Other Eukaryotes - 633 (source: NCBI BLink).	SNGVINQLCRNGRFKEAIDILCEKNLVKEAVQLLN HIQHPSASIYSSLLLICLQQRSLNEGRRVHDHIKA SGFVPGVFISNRLMDMYVRCGSLSDAQKVFDEM PERDLCSWNTMIKGYGKTGKVELARKLFDEMPQ RDGFSWTAMISCYVQHRRPKEALNLYRKMQSC GSWDSNKFTISSVIAASASLPSLLSGKETHGRIM RLGLECDSAVWSALSDMYAKCGCIEGARHIFDKT VDRDVVSWTTMIDGYFEDGRWKDGFGLLSDML RSRVRPNEFTFAGVLNACADRALENVGKQAHGH MIRIGVDSDSFAASALVHMYSKCGNISSAKNAFE GQCTPDLVSWTSLIVGFAQNGHTEESLELFELML RSDTEPDHVTFVGVLSACTHAGLVEKGIEYFHSI KDRHGLTPIVEHYSCLIDLLSRAGRFKEAEEYIMN MPMKPDKFLWASLLGGCRIHKNVELAELAAKALF EIEPENAATYATLANIYATAGRWEEVAKVRKMMD NRGVVKKPGLSWIEVRRRVHRFLVGDTSHLSYN EIHAFLAEISKKMKQEGYVPDTDFVLHDVEEEQK EQNLSYHSEKLAVAFGIVSTPPGTPIKVFKNLRTC VDCHTAIKFISRITKRKIIVRDSNRFHFFEDGSCSC GDYW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
221	sp_092890_sdyj	#N/A	#N/A	0.833536332	MATE efflux family protein; FUNCTIONS IN: antiporter activity; drug transmembrane transporter activity; INVOLVED IN: drug transmembrane transport, transmembrane transport; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Multi antimicrobial extrusion protein MatE (InterPro:IPR002528); BEST Arabidopsis thaliana protein match is: MATE efflux family protein (TAIR:AT4G39030.1); Has 7799 Blast hits to 7783 proteins in 1733 species: Archae - 152; Bacteria - 6456; Metazoa - 30; Fungi - 49; Plants - 387; Viruses - 0; Other Eukaryotes - 725 (source: NCBI BLink).	MQFKAIVSSSPYIQFKLPIINQNSLSSHFRRPKTS PKPFNLCISKFRPNPRNLRKNCCISPSQEYLGEN LETHEPESEISGFEINPENQNLQFPKPTNEVDDF ANQSTWDQIFEIIKFSGPATGLWICGPLMSLIDTA VIGQGSSLELAALGPATVLCDYVGFVFMFLSIATS NMVATSLAQKDEDQVQHQISMLLFMGLASGVLL LLLTKCFGEWALTGFVGVHNVHLVSAANSYIQIR ALAWPAILVGSVAQSASLGMKDSMGPLKALLIAS ALNGMGDLVLCTFLGYGIAGAAWATAASQIVAAY AMVEALNRKGYNGYSFSVPSLSELQEIFGIAAPV FLTMVSKVTFYGLLIYFATSMGTHVVAAHQVLLNI FGMCSVWGEPLSQTAQSFMPELLYGANRSLVK ARSLLKSLLIIGVSVGLLLGTAATAVPFLVPNFFTS DTVVIGEMRNVWICFFLALSVTPCILPCEGTLLAG RDLKFISISMSGCFAFGALVLLFVRSQQLGLAGC WSALVGFQWARFSISLSRLLSPSGVLFSEDLDRF RLADLKVA
222	sp_094570_sqpg	#N/A	#N/A	0.968932	CLAVATA3/ESR-RELATED 26 (CLE26); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MTTTTTTKSQDKHNVKHEVTTVGRGKLSARSKV DINYMSKRKVPNGPDPIHNRRAGNSRRPPGQA
223	sp_095520_qhxy	#N/A	#N/A	0.93105	uncharacterized protein	MAFHSSCSSCGSNLPEFPQQLNLIRSPPVPPTKP VHTLILSTNWTGVTTTTPPSFTFPGVNINQTIYSGI RNQGELYKISTFTLRDFLMNPLEVPQSLWNLYMV VPGDDGPKTLRPAHIPSLQDVDSMSKVAGIVLR WWTQRLAACSVDRLVIPAELQFLSDPEPVLPGK LYRLLITEDWVRAQYELSCNNVTLFTCAHSDKKV NRQISLSCAWGWKKGSEKLMGGLKLRDNVYEM YSPYLERFVTKDNISFDARVETNSLPKLNRLMVIM LWNSRGVARRWFRRDFDQLMDDHSPDIVVLTD TRVSKRNAEEILMDIPEFDSVSVVEALGFSGGIAIL WNSLHYTFIRIGKEQRALHGVIQGAESYD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
224	sp_095840_iuti	#N/A	#N/A	0.833641383	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT3G29230.1); Has 37071 Blast hits to 13818 proteins in 252 species: Archae - 0; Bacteria - 21; Metazoa - 136; Fungi - 146; Plants - 36104; Viruses - 0; Other Eukaryotes - 664 (source: NCBI BLink).	MMNTKLLNLSALFSSYKLSYDEVCSILQRCTKAK ASYSAKKIHGMLIISGTDVSRMSLDAKILGVYASC GNLRYAHQLFDKMPNPNVFAFNWMISVMSFHGN YQEAIGYFSLMQKMGTLPNKYTLSFVLKSCVGLL DLNKGRGVHAVMNKMGFECDALVCNALVDMYC KCGNTQYARRLFDRMSSKDIASWTSMISGYCNT GKLNEAIALFEQMKLEGFEPNDFTWNAIITGYAQ SGDRDGALELFSRMNREGLCCDLITWNALISGFS RSKHPFEALKLFREMLIAGIRPNHVTVTGLLPVCG LIGSPQRGRELHGLICRLSLEVNVYVASALIDMYS KSGSVKAAKNVFEGTLVRNVALWNVIIGCLGKNG VVFESLRLFEEMQEQGFQPNEVTLVCVLSACSH GGLVEKAMEIFRSMKEKYQVEANKEHYACVVDL LSRAGEVEEAYELIKRMPMEATESIFGAFFNGCNI HGRRDLAEKMADSLLKAAWKKPGALVTLSNIHA AGEEWQGVQNVRKLMKGIGVYKKPGFSL	
225	sp_101590_doje	#N/A	#N/A	0.85540123	Ankyrin repeat family protein; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT5G35810.1); Has 10613 Blast hits to 6237 proteins in 350 species: Archae - 27; Bacteria - 619; Metazoa - 5584; Fungi - 682; Plants - 1506; Viruses - 31; Other Eukaryotes - 2164 (source: NCBI BLink).	MSGQQPPSGSQVSQQASSSSSTTHDVTLSIIPSP TSSSATVPPPPRPPPPPLSAAVISSSRADFLKFGL PLYKAALEGNWPAAEQILQQSPDWIRHPIAKGGE RVLHIAAAAKRTHFVKELVKLMIEGNIDDLALSND AKNTALCLAATSGVVEIAKVMVDNNGALPNIRGS QDMTPLQMAVLLGRRDMVWYLIGVTDNNQLTDV DRIAILTSSINTNLFDVANHILKEHPELALLRDGKR ETALHALARKPVRSGDYHEQSIWKTLVGLGGKL AQGKFPQQSLELQLVQRLWDEVIKQEDKNISRLI GSPWRLLFIAAKCGNVEFLTTLIRTYPDLIWKVDE ENRSIFHIAIINRHEEIFKLIYELGAIKDLIAVDKDGN DNNMLHLAGALAPSHRLNCVSGAALQMQRELL WFKAVEKVVSSEYAEAENKSYKTPRALFTERHE GLRAKGEEWMKQTAQSCSVVAALIATVVFAAAF TLPGGIDSKGTPVFVNRRLFIVFAICNALSLFSSC AAILMFLSILTSRYAEHDFLKSLPLKLMLGILTGMFI SIVTMMAAFTVTSFITFREGVKSVNWIIPIALFAA VPVALFAFCQQYPLLLDIYKSTYRSWSLFQPTKPK LFPSYSS	
226	sp_114190_itms	#N/A	#N/A	0.9455233	Polyketide cyclase/dehydrase and lipid transport superfamily protein; CONTAINS InterPro DOMAIN/s: Lipid-binding START (InterPro:IPR002913); Has 197 Blast hits to 196 proteins in 72 species: Archae - 0; Bacteria - 22; Metazoa - 96; Fungi - 2; Plants - 31; Viruses - 0; Other Eukaryotes - 46 (source: NCBI BLink).	IYQQWDPDLVEARYIKDLEENLSIIRLRFGDASRP LFKNREFIVYERRETMDDGTLVVAVASLPKEIAAG LYPKQNKAIRGLLLQSGWVVVEKLEDHSCMVTYV VQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
227	sp_116820_oodk	#N/A	#N/A	0.994734	RNA binding;RNA-directed DNA polymerases; FUNCTIONS IN: RNA binding, RNA-directed DNA polymerase activity; INVOLVED IN: RNA-dependent DNA replication; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: RNA-directed DNA polymerase (reverse transcriptase) (InterPro:IPR000477), Protein of unknown function DUF1204 (InterPro:IPR009596), RNA-directed DNA polymerase (reverse transcriptase), related (InterPro:IPR015706); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF1204) (TAIR:AT4G20700.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MWASRPNMSSSSVSVEADDADLYFDGHGSVQP GLLSSADSPFQAQFLSRPCGILSSGPTFPDPFCV DQNNCSGSPMPLSPNPNPGRRPPFGPMEAGPSN WVEYPPDYHLSPVPDPGSPSPSLSESFPVTPAT DSGYFPDSSSNADFDFAGNNGDSSDAESEVVQP LLAAIPPADGVLAVDTSPVYSPASVDSSSTRLQP SLPESVSAAFAVSNSVDHPESVTQFRPIGLCNVIY KCIAKCLTHRLQGVMSSLVSETQNAFVPGRLMS NDCLLAHELISFVNNCPRSKSKCYAAIKLDMNKAY DRVLWDFLFKTLSAFGFPPYWVHIIRQCVSTVSY QVLVNGEPTSSFRPCCGLRQGDPLSPYLFVLCM ENRRAGPSIAAPEARSRSRLQKLVAEDELPKQSI LQTQASMVETDPSWPEPKLPEVRDKSPPRSRAYC KRSKLIDRSCLQPQAQCRSSRARPKLSSSMMKLI AEDPVSEQKQMMISPGGSCRGFLQKVLTKSSVS SSSPVSKALRAVPSSAVLKPKAEVVAASQGPSR V	
228	sp_118780_oaqp	#N/A	#N/A	0.995719	uncharacterized protein	MEERAQDSPSPPQIQSNLSALSQRFTPQQLLTAS AVMKAMAELSSRRTEGNQIVVTGNRPGAMPVRN LYETLEHEVVLPAQPEPILVQSENPGRRKRHSSR RRERSTRRRESVSEQEGSFPAGRESTPYHEEYI VQSPQP	
229	sp_119720_znuk	#N/A	#N/A	0.992838	uncharacterized protein	MVASLPLCAAILSNKNNKYLSYVKEGKTHKHLRF SGEEIVSRSTKFAIERSSKEDDENTELVHLRCCH NNKYLVKKSEAEGWIIAAAEKPEEDESVWSSTLF EPIIVESHDGYGRTVLRFKHVQSQLYASLSKDDD DDDDDDDDSLSKCLQLRQEIEDVINSSDVVTIID WGTLVILPKRVAFKSSSDGKYLVDDQRSSAHPCI RSSVDKKDDARVIHEVYTNPDGTLSIKSSDADGF WTSHEGGEIDVEPFADNTSEQKLATMEIKKGIFQ TFVPIKISSNEIALRGLPLNKFCLIDGTSKELVCKA NSILRETKFVVEEVVISREVSNIEYHLDDARIYDEA TVVLSTNIIHNTSENAFTTEYSFRFVDKRISRWTN GGSLGLGVPVNIKSAGIPLITENGIVISGEFTGTYE WGESKTETNEKEMTYNTSVPPMTQATIRLVSTS GTCDVPFSYSVREVHVDGKTVVYTMDDGLYTGI NFYKLQTQMETKPI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
230	sp_120350_rpma	#N/A	#N/A	0.888888	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, kinase activity, ATP binding; INVOLVED IN: transmembrane receptor protein tyrosine kinase signaling pathway, protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat protein kinase family protein (TAIR:AT3G47580.1); Has 213556 Blast hits to 133070 proteins in 4776 species: Archae - 194; Bacteria - 22980; Metazoa - 68117; Fungi - 10236; Plants - 86850; Viruses - 328; Other Eukaryotes - 24851 (source: NCBI BLink).	MSCSNSFFYWIIFSLSVFFCLEFSIASDTNTDKIAL LEIKAKISDDPLGVTSSWNETLHFCEWYGVTCG GRHQRVTVLDLQSSKLTGTLSPYLGNLSFLRILDL RNNSFSSVLPPEIGRLHRLQTLQLSNNSIGGEIPS SISNCYSLTGIYLILNKFVGEIPPTVGSLSHLQFLK LTRNNLTGSIPSSIGNLSSLSIFSVGGNNLVGKIPD SLGKLKNLTSLTLCDNKFSGELPPSIFNLSLLTYL CLYQNDFEGNLPPDIGNTLPLLQWFSIFNNRFSG HIPTSISNASNLQVLDLSKNNLRGQVPSLHKLVNL IHLVLYSNSLGFGQAQDLNFVSSLANATNLQWLE IDQNKFKGVFPKIICKFSSITKLVLSDNNIAGEIPNC IENLAEMQMFDVSKNALSSVIPHSIGKLQNLETLN LGDNKLTGVIPPSIGNLTKLSMTDLSNNRLEGQIP PALENCRSLIGLFLSNNNLSGSIPSQLYNIPSLSIA LYVDGNHLTGSLSEEVGQLKNLEALDVSGNMLS GEIPSSLGSCLSLELLGMDENNFQGPIPNALQKL TGLVELGLSYNNLSGKIPTFLASLNLQILNLSYNNL EGEVPIGGVFNNATNLSIIGNKRICGGISALKLPHC NVSPTVQRRKLKNRKKLILVILSTILGVTLLVLLV LSYTCCLEKKTKETTTSNDSEKNRNFSYQTLFKA TNGFSSENLIGTGSSGVVYKGILEDGSTVAIKTFN LEYRGASKSFLVECSVLRSIRHRNLLKVITVCSSI DYQGKDFKALVYEYMVNGSLDNWLHPIEATRGV EDTNDDSRNLNLRQRLDIAVDIAFSLEYLHHHHG ASIVHCDLKPSNVLLDENMVAHVGDFGLAKFLMK DTDSHAKQSNSIGVRGTIGYTPPEYGAGNEVSTR GDVYSFGILLLEMFTGKRPTNEMFKGGLSLHCFV KEALPEHVIKILDHVLAEDINFEETDNSMMLEALIS ILGIALCCSSEVPRERLNMNDVTAKLSSIRNRLLG TCSNKRGEFKQQHEGDINV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
231	sp_121650_zwed	#N/A	#N/A	0.942838	Oxidoreductase family protein; FUNCTIONS IN: oxidoreductase activity, binding, catalytic activity; INVOLVED IN: oxidation reduction, metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oxidoreductase, N-terminal (InterPro:IPR000683), Oxidoreductase, C-terminal (InterPro:IPR004104), NAD(P)-binding domain (InterPro:IPR016040); BEST Arabidopsis thaliana protein match is: Glyceraldehyde-3-phosphate dehydrogenase-like family protein (TAIR:AT1G34200.1); Has 17148 Blast hits to 17147 proteins in 2160 species: Archae - 344; Bacteria - 12203; Metazoa - 250; Fungi - 605; Plants - 99; Viruses - 0; Other Eukaryotes - 3647 (source: NCBI BLink).	MENRRKVKVHAIFTFLADDYFLKNDIRVKPDLDAL GALGDVGWYCIRAILWAANYELPKFVTALRNPIR NDDGVLLSCDAALHWEDGKIATFHCSFLSHLTM DITVIGTNGTLHVNDFLIPYKASEVSFKTASKCWF TDKALEWTHKSSEQVVSTDLPQEVLMVKEFSGL VDGIKFRDLAPEMKWPTISRKTQLVIDAVKVSIDK GLQAVEVVY
232	sp_123780_pgiy	#N/A	#N/A	0.936511	O-acyltransferase (WSD1-like) family protein; CONTAINS InterPro DOMAIN/s: O-acyltransferase, WSD1, C-terminal (InterPro:IPR009721), O-acyltransferase, WSD1, N-terminal (InterPro:IPR004255); BEST Arabidopsis thaliana protein match is: O-acyltransferase (WSD1-like) family protein (TAIR:AT2G38995.2); Has 1220 Blast hits to 1207 proteins in 180 species: Archae - 4; Bacteria - 957; Metazoa - 16; Fungi - 0; Plants - 224; Viruses - 0; Other Eukaryotes - 19 (source: NCBI BLink).	MTPNLQIVTNGGKPENDEAEPVSPTGQYFNSKV LSVCVLAILEIDVPIDDSCVIPQLRDVFLPMNPRFS SIMISDNKDVKQWKRVEVNLQDHVVVPSVPDGL SVESYDKYFDEYLTKITVDPLPQDRPLWELHVIKY PTSKAAGHFIWKLHHALGIDGYTLMGVLLSGVNR ADDPSLPLTFPSTRSSSLVTNNKMNIISWVPRTF SAIYNGVVNFGWSFLKSTCKADDKTPIRSGNEGL GFHPMKISTIELSLDQIKFIKTKLGATVNDILAGIIFL GVRKYMQATDTESGNSESTALVLFNTRNIGGYM TAEQMKKAQMKIWGNQFAFLHIAIPQLINDKCSN PLDVYEARKQISRFKSSPSVYLTAQCLELLRKC KGPEAAAEFIQSTTNKASILMTNMIGPIERVTMVD HPVKGMHFMVTGSPQDRSRD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
233	sp_127590_adcq	#N/A	#N/A	0.8368773	Oxidoreductase family protein; FUNCTIONS IN: oxidoreductase activity, binding, catalytic activity; INVOLVED IN: oxidation reduction, metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oxidoreductase, N-terminal (InterPro:IPR000683), Oxidoreductase, C-terminal (InterPro:IPR004104), NAD(P)-binding domain (InterPro:IPR016040); BEST Arabidopsis thaliana protein match is: Glyceraldehyde-3-phosphate dehydrogenase-like family protein (TAIR:AT1G34200.1); Has 17148 Blast hits to 17147 proteins in 2160 species: Archae - 344; Bacteria - 12203; Metazoa - 250; Fungi - 605; Plants - 99; Viruses - 0; Other Eukaryotes - 3647 (source: NCBI BLink).	MAESGSNAIRFGILGCAEISRKVARAISLSPNSPP HAIGSRSLEKAQTFATTNSFDPSTKVYGSYDEVI DDPEVDAVYYPLPTSLHLQWVVRAAEKKKHVIVE KPPALDVKELDVMLNACKANGVQYMDGTMWMH HPRTKIMTDFLADSQRFGHLKA
234	sp_129340_kdia	#N/A	#N/A	0.8335515	uncharacterized protein	MLAASAVIDAKYEARTTASIMTEELLKAEVKAMSS LREVERAVASKLKQNVMMTEDKCSPSGNICSGF GPPEQCCSGACVPHPILRIFVCQ
235	sp_131200_inri	#N/A	#N/A	0.8358778	Ankyrin repeat family protein; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT5G54620.1); Has 21854 Blast hits to 10883 proteins in 482 species: Archae - 19; Bacteria - 1460; Metazoa - 11922; Fungi - 1177; Plants - 2352; Viruses - 52; Other Eukaryotes - 4872 (source: NCBI BLink).	MEKERDKSESVVTVLYEAALTGDVPSMLNLLEQ DQFILERCIIEKSCHFMQSPLHVAVNMGHIQFVRE ILRRKPQFAEVVDRLRRSSPLHIACAIKGRLEIVNI LVEEKPNMCFAQDQDGRTPIHVAVVNGNVEVME VLLKEKHQAAWERTMGGETVLHLCVKNNQLDAL KSLVNKTDDPELLNFADSNGDTALHLAVAYKESE MVECLLAQNKIAKDVRNKKEKTIVDIHKETNKGSE ADKKIDRLLKKYKAKPAKYGLKRQTENKWLEEQ NTALMVVAASIATMAFQVGMNPPGGVWQDNNN GHEAGTSIMSYDSDGADYNVVLVSNTIGLISSLSV ILLLISGLPCKRFFVFLRITLWISVSATAATYLLAI GYVTPNVINKASLVNDAMDYSVEVWLWLVAVILL GHGVLFILKLIGHDRRQKIREFYKTHFVKPVEAVP KRVLESQITCPIV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
236	sp_134450_jzcs	#N/A	#N/A	0.9333734	C2H2 and C2HC zinc fingers superfamily protein; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: response to chitin, regulation of transcription; LOCATED IN: intracellular; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), Zinc finger, C2H2-type (InterPro:IPR007087); BEST Arabidopsis thaliana protein match is: C2H2 and C2HC zinc fingers superfamily protein (TAIR:AT1G34370.1); Has 32454 Blast hits to 15503 proteins in 208 species: Archae - 0; Bacteria - 0; Metazoa - 30756; Fungi - 38; Plants - 696; Viruses - 0; Other Eukaryotes - 964 (source: NCBI BLink).	MIQGEVLCFTNGGDAAEGGNGDLSNLYNNGVLL ESSYDGVVVAGGSSSEENLSIRSLLYSISMLKEK VGDVESFVSIYLNQQENHYYDHSHQTTVDQQLT SSMAFSSMANLIQEIVHTASFMMLSCEQMALSGT TTTTTTQIGNINNELHTMQKANNQEVFNNGDSFY HHHASKNTNQESPSSLSVEAAYDLVEVDAADLLA RYTHYCQVCGRGFKREANLRMHMRAHGDKYNK LSLMSNNNINNNNNSLKQNKNDNEVEHNNNTRT KAVAVAVAPPRKYSCPQKGCRWNKNHVKFQPL KSMVCVRNHYKRSHCPKMFVCKRCNMKEFSVL SDLKTHEKHCGDNVKKLYCSCGTSFTNNHKLMD HVALFLGHTPLIYP
237	sp_137300_ruhx	#N/A	#N/A	0.9334656	LRR and NB-ARC domains-containing disease resistance protein; FUNCTIONS IN: ATP binding; INVOLVED IN: apoptosis, defense response; CONTAINS InterPro DOMAIN/s: NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: NB-ARC domain-containing disease resistance protein (TAIR:AT3G14470.1); Has 38623 Blast hits to 22788 proteins in 858 species: Archae - 54; Bacteria - 2838; Metazoa - 8237; Fungi - 658; Plants - 25379; Viruses - 2; Other Eukaryotes - 1455 (source: NCBI BLink).	MPLNNLRGELKIQFRESYSYDATNAEEGSYLLSK KHIKALVIKGAKYETPIVLDERLLEVLQPHHDLILIKI YGYKSIRLPGWVAAIAANLPRLVRICLERFERLQH LPQILSQLQHLKILQLDNMPGVEYMEDDKGNIPCF PLLEELILWNFPKLKGWWRPEVGVETDGVIPGL HKLTIFGCPNLTSFPGIPHVKEKRANKFNPYW
238	sp_155920_gjdz	#N/A	#N/A	0.9337417	uncharacterized protein	MAKFYVEVIILKVIIGSILLSNPIESLCSEKLKPCIESI VSFNNNQSQISIIETGFNEGLALQMEANLLDQCC GRFTVETISRDKQCFCSSISSIQEHRLSNNITLLLS ECSTSDGMLLFYDDDLCQDIQAPSSAPVFPLPSY VLSHTQLSPTSTPTPIPTPPRNKKIIIAIIVPVTVLVV LIALCIGFFSMKRKEKPNVAQSHHQSGKTFK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
239	sp_156640_yxdd	#N/A	#N/A	0.8283156	NAD(P)-binding Rossmann-fold superfamily protein; FUNCTIONS IN: oxidoreductase activity, binding, catalytic activity; INVOLVED IN: oxidation reduction, metabolic process; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Short-chain dehydrogenase/reductase, conserved site (InterPro:IPR020904), NAD(P)-binding domain (InterPro:IPR016040), Glucose/ribitol dehydrogenase (InterPro:IPR002347), Short-chain dehydrogenase/reductase SDR (InterPro:IPR002198); BEST Arabidopsis thaliana protein match is: senescence-associated gene 13 (TAIR:AT2G29350.1); Has 121057 Blast hits to 120745 proteins in 3594 species: Archae - 991; Bacteria - 79410; Metazoa - 5524; Fungi - 6125; Plants - 2737; Viruses - 5; Other Eukaryotes - 26265 (source: NCBI BLink).	MASIVAVDVGSVYGPAKGAMNQLAKNLACEWAK DNIRNSVAPGVIWTAITEGLKNYKEYKEAFESRIA LGRVGNPDEISPVVAFLCLPAASYITGQTINTSA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
240	sp_159110_dayj	#N/A	#N/A	0.943618	P-loop containing nucleoside triphosphate hydrolases superfamily protein; BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT1G65780.1); Has 6257 Blast hits to 5371 proteins in 1056 species: Archae - 197; Bacteria - 1787; Metazoa - 1336; Fungi - 1200; Plants - 672; Viruses - 106; Other Eukaryotes - 959 (source: NCBI BLink).	MFFLGSDILTYLYAKQIKKVPDTFSSTDHYLSSFK HLLKEEIRAEFSAGVESIAQSPACEISKVRISKHHK PPKSLYYEIMKKDIREVEKNSEHNYMLQSGDLIVL TDMRPRRMEDLIREKDNLFLFAFVASSNVDEPWI QIFSSKEFEPKLTTKKHGRFFATFLMNMTTSFRI WSALNPDPKSSSMGLIQKVLQYDSTVDDDCSLC VNEKPFNVKGLNVFDLISTLGLDESQQKAVLNTIS MRNCSHQSNNVKLIWGPPGTGKTKTVASSLFVL LKLKCRTLTCAPTNIAVVQVANRLVRIFLESNGEL DTYGLGDIVLFGHEERMKINDHDELMDVFLDYRK EVLGECLEPLHGWKGTLESMKSFLEDPQTRYND YLQQHEVEDEENGSKKRIDGVKKNKVSDMVLKT EDLMIFEEFFKKRFSVLADRLVYCLKNMYTHLPTT SLPLNVAKQMIRLVHVLKTITNARKKVSDHKNLM MKREEILCILNYLSDQFPKLKFKGSVHEFCLSNAR LLFCTASASIKVSKNVEMVIIDEAAQLKECESMIPL QVPGLKNAVLIGDDRQLPAMVQSKALGDLNFGR SLFQRLAKIGKKKKHLLNIQYRMHPSISLFPNKEFY ENRIIDAPKVEEISYSRSFLKGAMYGSYSFINVSR GKEDFEKGHSPRNIEEAVVVDHIVAKLFKGLDFK KQKLTVGVISPYKGQVCLIQEKIGNKYANYKENFA VNVRSVDGFQGGEEDVIIISTVRCNGNGSVGFLS NHQRTNVALTRARYCLWIVGNWTTLAKSGSVW KKVAIDAKTRGCFYDADDDLHSKQAKTGSLAKIN QLSVSKPDFFGSLNLGKAKWKVFFADAFKSSITSI SSKKCQNKAQKKLKKIADGWRLQSNTKEAEKCV ATCGIAPELLELYKVKEQLYLVWSVDIIQEESKYT QAIKVWDVLPASKIPKLAKELDILFSKYTSEFMNR CKYKSIEGDLVLPMSWQLHSVLSAENLSDQFSV MNIRDDRQESSTGSLRDNMKAGSGISSSSRRRR RAGTA
241	sp_161290_shyy	#N/A	#N/A	0.943823	uncharacterized protein	MESLRSFLCFKTSQSSVVFPTPTTTTFTKLPSRF WHKAKAGYAVFGAKQAGKRKGLVSLSKDKELCE GYADIETMWKMIHSTCSKDIHSDGNCKRPRYLVF CFKPS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
242	sp_170040_wkpp	#N/A	#N/A	0.934632	Major facilitator superfamily protein; FUNCTIONS IN: transporter activity; INVOLVED IN: oligopeptide transport; LOCATED IN: plasma membrane, membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oligopeptide transporter (InterPro:IPR000109), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT1G52190.1); Has 5584 Blast hits to 5329 proteins in 963 species: Archae - 0; Bacteria - 2098; Metazoa - 610; Fungi - 414; Plants - 2170; Viruses - 0; Other Eukaryotes - 292 (source: NCBI BLink).	MAIEEEEKMAFLLDHQNSEHENPKPRTKGGLITM PFIIVNEAFEKVSSYGLLPNMIFYLMNVYHLEAAK GSIILSLWSASSNAMAIFGGFLSDSYFGRFRVIAF GTFCSLLGTTMIWSTSMIPWMKPSPCAMSSQDC DSPTSLHYLVLFGSFGLISVGAGCVRPCSLAFGA DQVNDKENPNNERVLDSFFNWYYASTGLSTVIAL TFIVYIQDKFGWQIGFGVPVFLMVISLATFLLGSPL YVHVKARDNLFIGLFQVPVAAFQRRKISFRAGAE VRYYCGNGSEATTPSDHIRCLNRACVIINPGTDF NPDGSASNPWTLCTVEQVERLKSIVRILPMWTT GFMMLVALNNSYTTLQAKTLDRHIFSSSFQIPAG TFTVFSIVTITFWVAFYDRAVVPLMAKYFNMPRG LGPRVRMGIGLLLSCVSLLVAGVVESIRRKVAIDS GLEDKPNEVLEMSAMWLVPQLVLLGLAEGFNAV GQIEFYYAHLYKGMAVSSLVGSLLVDVVDSFSSV GGKVSWLSSNLNKGHVDYYYWLLAFLCLVNFLY FLLCCRMYGTSESHSESHRSRRYESVEHESA
243	sp_170860_dkxg	#N/A	#N/A	0.934144	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, kinase activity, ATP binding; INVOLVED IN: transmembrane receptor protein tyrosine kinase signaling pathway, protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat protein kinase family protein (TAIR:AT3G47580.1); Has 213556 Blast hits to 133070 proteins in 4776 species: Archae - 194; Bacteria - 22980; Metazoa - 68117; Fungi - 10236; Plants - 86850; Viruses - 328; Other Eukaryotes - 24851 (source: NCBI BLink).	MEISNQVSLVKLDVSSNKFSGEIPDVFNNLPTLQ KLYMAHNFFHGLIPQSFASLKSLLEVDLSHNRLS GPVPRYLSSFPLIYLNLSYNNFEGEVPTKGIFANA SAISLVGNNRLCGGISELHLARCVKKAPKKTRMY YTLKLVIFISTIIGILATGSCIWLFLTRHYKKREPVS LDAIGKESFSKVSYHMLLKATNGFSMGNLLGSGT FGSVFQGNLDGKTVAVKVLNLQQRGGNKSFMA ECEAMRNIRHRNLVGIITAQLAQVWISKEKISKLL SHALNYLHHECGNSIVHCDLKPNNILLHHDMVAR VSDFGLAKVLARPLHPNQSSSVAVRGTIGYVAPE YGLGAEASPKADIYSYGILLLELMTSKRPTDNMF MEENLHMYSKDALHDRVLKIVDSTLLEHENEEVE SSEIKTQVVLQNREECITLVVKIGVACSNQLPRDR MKISDVIIELQEARKILLISKQRQTFSTITQM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
244	sp_172940_kgac	#N/A	#N/A	0.8821345	DEFECTIVE ANTHER DEHISCENCE 1 (DAD1); CONTAINS InterPro DOMAIN/s: Lipase, class 3 (InterPro:IPR002921); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT4G16820.1); Has 1715 Blast hits to 1709 proteins in 324 species: Archae - 0; Bacteria - 331; Metazoa - 67; Fungi - 376; Plants - 633; Viruses - 9; Other Eukaryotes - 299 (source: NCBI BLink).	MRLSMKSLVLPYDLPMSSGTPYVPVPVPVPGHQ QCSITKRTQELTFQSQQSINRQSLNQTKLTPVVR QSRAANRWKEYMGVRDWDGLLDPLDDHLRDEI LRYGSFIEATYRGFEENPHSSSYGSSKYKKKSFF EQCGLPTPGYRLTRHLRATSGIQVPGWASSSWS PMQTSWIGYVAVCTNKAMIKQLGRREVVIALRGT ATCLEWLENLRATLAPVPGPISVENDNVGPWNV DSPPMVESGFLSLYTSGNDMWPSLQDAIRIEIGRI LDLYQDKEPLSITITGHSLGAALATLAAYDIKTTFD RAPLVTVISFGGPRVGNQSFRSLVDKQGTKILRV VNPSDLITKLPGFVTDNNEDFVNIDTSPSATKKIGI KNSPRTSHNNNKNNNRSNRPGTITSVSDSTRF GKILSSHFGKGKIKSSSKSRLGVTKSGSSKMKLM GNILMNWVQQFRVDDTRWWVYAEVGRELRLEN RESAEQLLQYYLSNMDVAKCHDLKTYLNLVEDC PLKNFVRRSLNKLTQRCTRNAHEKDNNINNNSG VLAL
245	sp_173100_ziaf	#N/A	#N/A	0.9346418	uncharacterized protein	MEDGFNRNVNDFVT
246	sp_176550_rgfx	#N/A	#N/A	0.9483444	Leucine-rich repeat transmembrane protein kinase; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Malectin/receptor-like protein kinase (InterPro:IPR021720), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat transmembrane protein kinase (TAIR:AT1G29740.1); Has 108967 Blast hits to 107923 proteins in 3166 species: Archae - 87; Bacteria - 12310; Metazoa - 39616; Fungi - 9390; Plants - 31491; Viruses - 414; Other Eukaryotes - 15659 (source: NCBI BLink)	MELVDPMLGTDYNTEQVLRMIKVALLCTNPSPAV RPVMSEVLRMLEGQMLVKELTLAPGLCNNDWFY EASKRYQDSTSDSETQTLIRFPNARNNANSLSDL SPIPTNHPKPPVIEPVQT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
247	sp_183980_psck	#N/A	#N/A	0.8573774	NAD(P)-binding Rossmann-fold superfamily protein; FUNCTIONS IN: oxidoreducase activity, binding, catalytic activity; INVOLVED IN: oxidation reduction, metabolic process; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: NAD(P)-binding domain (InterPro:IPR016040), Glucose/ribitol dehydrogenase (InterPro:IPR002347), Short-chain dehydrogenase/reductase SDR (InterPro:IPR002198); BEST Arabidopsis thaliana protein match is: NAD(P)-binding Rossmann-fold superfamily protein (TAIR:AT5G51030.1); Has 50499 Blast hits to 50443 proteins in 2986 species: Archae - 509; Bacteria - 33822; Metazoa - 3124; Fungi - 3071; Plants - 1791; Viruses - 1; Other Eukaryotes - 8181 (source: NCBI BLink).	MMKMKQFNLIISRKIIPTLFYSTASFNREASVSLV QGASRGIGLEFVKQLLVRDKKGHVLATCRNPDQ STGLLELKNQFSDRLNLLKIDVTDENTIEAATTSV RETYGYLNLLVNTSGILSIPNLLHPETTLAKVQKS ALMLTYEVNAVGPLLMIKHMWPLLKEGGGKGTM RDAAVVANLSARVGSIGDNNLGGWHSYRASKTA LNQLTKTVSVELARKKDPVICLLLHPGTVDTDLSR PFQRNVPQDKLFTREFSVSKLLSIIDGAKIHDNGK FFAWDGQEIPW
248	sp_187280_gxza	#N/A	#N/A	0.8334431	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G17580.1); Has 40 Blast hits to 40 proteins in 11 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 40; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAASWVKTITSPFKKACTFVKHQPPKDKKSKQE NENTVMALHGEVMACTYEDVQVMWSMLDKSKP GACNMTS
249	sp_188860_zcgg	#N/A	#N/A	0.8334132	unknown protein; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF247, plant (InterPro:IPR004158), Protein of unknown function DUF862, eukaryotic (InterPro:IPR008580); BEST Arabidopsis thaliana protein match is: Plant protein of unknown function (DUF247) (TAIR:AT5G11290.1); Has 1967 Blast hits to 1844 proteins in 183 species: Archae - 0; Bacteria - 6; Metazoa - 223; Fungi - 83; Plants - 1477; Viruses - 0; Other Eukaryotes - 178 (source: NCBI BLink).	MGDLTKWVSNSISEALELLPAVSSYCGTYRVPEN LRKVNEDAYRPLLVSIGPSYYGDPRLVAMHQQKL RYLQSFLQRYTCYNLDYYVQVIREWEVVARQCY VESLDTTRLTSPQFIEMLLVDAAFLIELFLRNNFHE FIEENDRIFNKPRMMLEVTRDIRLEENQLPFFILKG LYDLVIQPSSQNKQVSFLDITYKFLTGKNEAVPAR IIDADVKHLVDILRLCYLPCSVRGQHPRTNPKFEF ASSVSELSEAGVKFVASPSENLLDIRFVKGYLEV PRFVVTDDTESLFRNIMVFEQCHYYFDSYIIDYFA FMDSLINTPKDVGILVQSGVIENWLGNNEEVANL FSNIFKQTRLKGSDFYYSGICKDLNAYAGTPWNR WKAILKHDYFNHPWAAISVVYAIVMLILTVLQVVT GFK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
250	sp_191130_rcps	#N/A	#N/A	0.998632	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity; INVOLVED IN: biosynthetic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: hypocotyl, root; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT3G60290.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MNSTIEKFVSSWSDGKTVPENYVFPTGQRPAEH LYPKSKISVPLIDLGKANGPDRSEVVRQIMDASQ NLGAFQVINHGVSIELMDNARDMFKEFFNLPAEE KAKYCSWDTKQKFILYTSKKSSIFGGML
251	sp_192960_cgkm	#N/A	#N/A	0.998238	Plant protein of unknown function (DUF936); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF936, plant (InterPro:IPR010341); BEST Arabidopsis thaliana protein match is: Plant protein of unknown function (DUF936) (TAIR:AT3G14170.1); Has 718 Blast hits to 418 proteins in 99 species: Archae - 0; Bacteria - 149; Metazoa - 50; Fungi - 25; Plants - 280; Viruses - 0; Other Eukaryotes - 214 (source: NCBI BLink).	MATLVPGILLKLLQHMNTDVKVAGEHRSSLLQVV SIVPALAGGDLFQNQGFYLKVSDSSHATYVALPD EHDDLIRSDKIQLGQYIHVERLEAASPVPILRGVR LVPGRHACVGTPEDIVATHSLGFLNNSSPGSKPT EKSKSAAKSLFGRKEKSPPSKLNSFGKDDLVDKK LSITRSKSISSKASVNSLDRKEALGRLKTSSSQSI PSSPTSCYSLPTSFEKFSNGVKHQAKIKSDRGTP KLGILMEKVSAKLGLVEKASSARSSSFITKKQPSV STPKKFAQVIDMGPKALRRSWEGSMETKSRDSL KVRTSKPDVISASRSTSVPKKNPADDRSSTKVES KGQKTSKISKVEAKDQISARTLPTNGDDLDRSSK PRISGRKSVDNGSNSSNGFPSNLAKVSASSRKLI DASVLWTSLPSSVTKLGKEVLKLRDSAQMAAIEA MQEASAAECILRCLSKYSEVSSSSAKEDDPQPAVE QFLALHSSLKSALVTVNSLSKMTSVGSSPDSQEA LSEETQKMVSERRKQAASWVNAALATDLSPFSV YSDRPTLTSLPTSSQSHNHKLNSGSPPFIVLENA SKNASPKTISKPRLSVSSKASSPGSGTPRRASDM LLSSQKARPLPPPPEWDRGNGLDEYMDLAEKLQ FESQNWFLGFVERFLDADVDSTALSDNGQIAGM LTQLKSVNDWLDEIGGGSAKDEEQEEEEEEVAR VSGETVDRLRKKIYEYLLTHVESAAAALGSGSQS LPPVRTTETKGRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
252	sp_193940_axoc	#N/A	#N/A	0.9838031	FUNCTIONS IN: molecular_function unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Autophagy-related protein 27 (InterPro:IPR018939); Has 138 Blast hits to 138 proteins in 57 species: Archae - 0; Bacteria - 0; Metazoa - 32; Fungi - 62; Plants - 33; Viruses - 0; Other Eukaryotes - 11 (source: NCBI BLink).	MGSHYSSKLTLFLLSILVGVNFSSVFAATCEYGT RERNKLYSYSLTSPVKNFRHGVLSEDGFYKVAA NDTIIWFQLCDGMIFNHDPPRCADCWDCGGPSR CGMGCSALVAKYTGGYYACTTIGREMNFDINLTD RKTPDKGVTITMSSYTDGVNCSLSVSVICDSKGIK EPESLKKLGACHYATELRHPSGCPKIAVHATGW GWFRLLLTILICSVGAYLLIGGVYRFFHLGIRGLDV IPNLDFWISLPYSIKSVFMSLVQRFRGPTYHHTSS YSSVNF
253	sp_195460_jgcs	#N/A	#N/A	0.9416336	uncharacterized protein	MYSSQVKPGGGLKINATDLGAPVPAATKSVCCG S
254	sp_209810_wuzr	#N/A	#N/A	0.927467	beta glucosidase 11 (BGLU11); FUNCTIONS IN: cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: carbohydrate metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 1 (InterPro:IPR001360), Glycoside hydrolase, family 1, active site (InterPro:IPR018120), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: beta glucosidase 10 (TAIR:AT4G27830.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	IEPHVTLLHIDVPQALVDDYESFLNRRIIEDFTAYV DVCFREFGDRVRYWTTLNEPNIFVFGGYDIGNYP PGRCSSPFGNCSRGNSETEPYIAAHLLLLAHGSA VKLYKEKYQAKQLGFVGMNLLSYHFSPITNTRENI IATRRTYDFLIGWFMHPLTYGDYPEIMKKNYGSRI PVFTEKESNLVKGSFDFIGINYYSIMETKDYSTGL SKYPRDFFADIGVKWTCMI
255	sp_211820_gknm	#N/A	#N/A	0.8442394	uncharacterized protein	MGLQNNSRKEGPEQETNFSENLSWKYFAA
256	sp_212470_xogd	#N/A	#N/A	0.9931177	BCL-2-associated athanogene 3 (BAG3); INVOLVED IN: regulation of apoptosis, apoptosis; LOCATED IN: cellular_component unknown; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: Apoptosis regulator, Bcl-2 protein, BAG (InterPro:IPR003103), Ubiquitin supergroup (InterPro:IPR019955); BEST Arabidopsis thaliana protein match is: BCL-2-associated athanogene 2 (TAIR:AT5G62100.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGGLEWEMRPGGMFVQKRNLDSNRNYSIPIPTI KVLVKFNSSYHEIHINSHASFGELKKRLAQVTGLH PQDQKLIFKEKERESRSFLDVSGVKDGSRIILVQD IINQENRFIESRRNSNIERAMKSVAEVSIEVDKLAC QVTSMESVICKGGRVADKELLTLIELLMAQMIKLD GIIADGDVKLKRKTQVVRIQKYIDSLDRMKAQNKV ASRYGGGQVGLVVSQQQQRPFIMQQQREQKNF SNGLMLTSKWEKY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
257	sp_002450_dsmi	0.932284	0.92356	#N/A	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; CONTAINS InterPro DOMAIN/s: Metal-dependent phosphohydrolase, HD subdomain (InterPro:IPR006674), Ppx/GppA phosphatase (InterPro:IPR003695).	MVIPLESQNLFASVDMGTNTFKLLIIHLEPSSAKFT TIHNHSDAVALGRNSSPSSSSPSSPPLISPDSQF RAIESLKKFQEILRSHNVTRTRMVGTSAIREAGNR SIFLNEVRENLGFDFHVTVLSGVEEARLIYLGVLQ FLHVYDKRVLTIDIGGGSTEFVIGDHGEVVYADSL RLGHLTLTQNFGESDLVGLRGYVRELIEGSGLVD KVKILGFEVAVGSSGTIRAVENAIWCSYGCDDRE KGGMRVLEEGSRRNLWWFSREELRDLVERLCT VEFSEREKVRRDEFFGRRSEFIVAGAVLLEEIMEL LGISEMQVSKYALGEGVVAEMVSGEFPEFDLNA NVRWKSIMRLSMRFNGNDRMRNGVQCSAIAKDI FEGLRKYTKVFEDHFTDDVCLDEKDFEYLEAACM LHNIGLYLGKKGYHKNSYSIVMEHSQLDGYTSEEI QLIAQLARHHRKKFPNFEYTSLQGFPREKFGILCS ILRLSVIIQKHKILKLGEIKFDDSSESFVLVLVKKGD QPILLDSKSLAKDIETKTRKEAEFFEGVFHQKLLV VVEMDGSESSARQ
258	sp_033270_jhzc	0.933559	0.952238	#N/A	REVERSAL OF THE DET PHENOTYPE 3 (TED3); FUNCTIONS IN: zinc ion binding; INVOLVED IN: protein import into peroxisome matrix, fatty acid beta-oxidation, peroxisome organization, photomorphogenesis; LOCATED IN: cytosol, peroxisome; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type, conserved site (InterPro:IPR017907), Pex, N-terminal (InterPro:IPR006845), Zinc finger, RING-type (InterPro:IPR001841); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MDRQTLASSSSPSSSISEVPSPDAWITTYKTLLP QWNSFPLSHQIIPITISRVNQFDAARLDVEMSAML KEQLVKVFSLMKPGILFQYEPELDAFLEFLIWRFS IWVDKPTPGNALMNLRYRDERVMEVRGKVRTGL EGPGLTAAQKMWYCIATVGGQYIWTRLQSFSAF RRWGDSEQRTLGQRAWLLLQHVEGFYKAASFS NLLLFLLTGKYRSLIERALRARLVYGSPNMNRSVS FEYMNRQLVWNEFSEMLLLLLPLLNSPLLKSLFR PFSKDKASESASNESSCPICQGSPSIPFIALPCRH RYCYYCLRTRCAATPSFRCARCSQPVTAMQRLG GSVDEKSSKK
259	sp_039080_ztzh	0.941236	0.949436	#N/A	cyclin-related; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; Has 5668 Blast hits to 3895 proteins in 309 species: Archae - 0; Bacteria - 250; Metazoa - 3167; Fungi - 631; Plants - 477; Viruses - 7; Other Eukaryotes - 1136 (source: NCBI BLink).	MSDSKGKVPRERSPSRLKSEVGDKDGDKKHMH QLQDALALESLPASDSKQEVKGASSMLDSKNNG KGERTGHSNTTAVPRRSYHQHDDRRTAARDGS RSGRRSTSERGWWKDSRGEDRAAHRSAPSDP KVREGKSRAEDKRGWWIDSRGEDRASHRSGPS DPKVREGKSGAEGGDSQVWRHDRFIEAEGDAQ PSKKRRPFREEKLPVKSEDVDNAVVPERSKVSH PTGGTARWEQKAHDRRQLERNEHDRAYLRERS FSNRGRVENTGFPSRDRFNGVSNNGKFIGRENF GGRKGYHSSGGRVEKWKHDLYDEANRSPTPKK EEDPVAKVEALLSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
260	sp_052000_anyg	0.943179	0.9538327	#N/A	Inositol 1,3,4-trisphosphate 5/6-kinase family protein; FUNCTIONS IN: inositol or phosphatidylinositol kinase activity, catalytic activity; INVOLVED IN: inositol and derivative phosphorylation; LOCATED IN: intracellular; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Inositol-tetrakisphosphate 1-kinase (InterPro:IPR017427), ATP-grasp fold (InterPro:IPR011761), Inositol 1, 3, 4-trisphosphate 5/6-kinase (InterPro:IPR008656); BEST Arabidopsis thaliana protein match is: inositol 1,3,4-trisphosphate 5/6-kinase family protein (TAIR:AT4G08170.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAEQNNEIISVGYALPSKKVDSFIKPSLVSYAKTR GIDFIPIDIDQPLLQQGPFHCVIHKVYGDDWNREL LDFKARNPNVPVIDSPAAIQRLHNRISMLDAVKNL KPASSDAALTPSVGTPNQILLSEEIMMRSLNSHQ MSVVFNNNGLKKVREFAPIVLQEFVNHGGVLFKI YVAGENVRCVQRKSLPDFNDSDHPFFLQFEEEG FMPFAQISNLATTEMQQEIDNIEMPPLDFVNNLAK GLREGTDINLFNFDLIRDSSVSGGNRYLIIDINYFP GFAKLPGFEEMLTDFFLDVVHRSSSAVEKA
261	sp_075130_figr	0.927784	0.9328463	#N/A	FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation; LOCATED IN: endomembrane system, ribosome, intracellular; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Ribosomal protein L34e (InterPro:IPR008195), Protein of unknown function DUF1666 (InterPro:IPR012870); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF1666) (TAIR:AT3G20260.1); Has 1252 Blast hits to 1091 proteins in 184 species: Archae - 4; Bacteria - 49; Metazoa - 367; Fungi - 151; Plants - 117; Viruses - 41; Other Eukaryotes - 523 (source: NCBI BLink).	MPCSRKQPLLKLLYNVTATFKFFFLFFYLFSLFIAK LFTYFFRTEEEEYEDDISYEDEDEDEEEEAAESE EEEEEEYRYSHMEEEEEEEEGDHLVADIVNDGEG LIFVANYGDDDYERNNDHSIVHSNSYETLFQDPV EDFNDEAEAGCFENSPALTSSSELFDRGGDYTH HGGDVGGDNYDVINEDYASRFESNHLYDHSFEL EEEEIHDKNDHENPHQECEAVYHDMPQNSPSYI EVKVKVAPSGNHDVAGEFDESKTDVVKQDSNFT RDEKFLIFAPPKSEAKKLTGLEMENEGIFEDTYTV GSTSKSSSEWRSSIRDSGTDDPFSSSSSRRSCPK WESYAVYQKYDEEMTFLDRISAQKLHETESLRSI QVEPRSISQRIVIVHKLGNNKEKKKSSGDDIYKNKTN PYNELEVAYVAQICLTWEALNWNYETFRRKTKQ KSKTELQDDDHGCPANIAQQFQQFQVLVQRYIE NEPYQRGKRPEIYARLRSAAPNLLQVPEYHDYE GEEKELVGGRIASALFLQILEDSIQTFMNFLKADR ENHCQVITALFKLRPRGSIDPTVLNVFKKCNKKKK LKLKEVERTRKCFKKRRLTREEEMEILMSLIDLKV VSRVLRMTEITQEQLHWCEQKMAKPTILEGKLHR DSSPLFFPAH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
262	sp_077670_jsex	0.941976	0.923094	#N/A	hydroxyproline-rich glycoprotein family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: stem; BEST Arabidopsis thaliana protein match is: hydroxyproline-rich glycoprotein family protein (TAIR:AT3G26910.1); Has 1387 Blast hits to 1097 proteins in 177 species: Archae - 4; Bacteria - 37; Metazoa - 593; Fungi - 281; Plants - 163; Viruses - 19; Other Eukaryotes - 290 (source: NCBI BLink).	MKLKDFERNVKDSHEDSFIREKRIIGQSAPLFTLQ KFKPAEKIRQLQTSSSRKFHSYALPKPGDTTRISA RSENRVLSTKPNPLLAENYHSPSLVPNTYEKLVR TEMHNLHA
263	sp_080320_atmm	0.936231	0.935044	#N/A	Mannose-binding lectin superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Mannose-binding lectin (InterPro:IPR001229); BEST Arabidopsis thaliana protein match is: Mannose-binding lectin superfamily protein (TAIR:AT5G38550.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAEKPATSNFVEGTKNAAQKVETAVTTQTEKISK SAGEMSEKTKNSVGDALNKAGEFVKPAPPPPEE SLMDKAKGLFK
264	sp_103930_jftx	0.942581	0.943846	#N/A	galacturonosyltransferase-like 2 (GATL2); FUNCTIONS IN: transferase activity, transferring hexosyl groups, polygalacturonate 4-alpha-galacturonosyltransferase activity, transferase activity, transferring glycosyl groups; INVOLVED IN: carbohydrate biosynthetic process; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 8 (InterPro:IPR002495); BEST Arabidopsis thaliana protein match is: Nucleotide-diphospho-sugar transferases superfamily protein (TAIR:AT1G19300.1); Has 1885 Blast hits to 1878 proteins in 434 species: Archae - 2; Bacteria - 935; Metazoa - 151; Fungi - 0; Plants - 716; Viruses - 0; Other Eukaryotes - 81 (source: NCBI BLink).	MPPTRPFLVFGVVVFGLLITIVSATATATIDGSQS KQRFREAPQFYNSEECPKIVSTEFEDDDHQDVSL PSSALDPDSVLVCSDEAVHVAMTLDLTYIRGSMA AILSVLQHSSCPQNIYFHFVISATRSNASELHTTIS SSFPYLRFDLYPFDDSAVAGLISTSIRSALDCPLN YARSYLANLLPLCVGRVVYLDSDVVVDDISTLY NFNLVDDKVLAAPEYCNANFTTYFTPTFWSNPSL SVTFADRHACYFNTGVMVIDLNRWRAGDYTRRI EEWIMELQKRMRIYELGSLPPFLLVFGGNIVPVDH RWNQHGLGGDNFRGLCRDLHPGPVSLLHWSGK GKPWARLDANRPCPLDALWAPYDLLQTPFALDS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
265	sp_112680_fkwa	0.941873	0.944838	#N/A	Carbohydrate-binding X8 domain superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: X8 (InterPro:IPR012946); BEST Arabidopsis thaliana protein match is: Carbohydrate-binding X8 domain superfamily protein (TAIR:AT2G04910.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAWNLLLPLLLLLLLLLVVPNISGDDNDGEDDPG QQWCVADVQTPDNELESAINWACGQQDVKCSM IQENQPCYQPNTTKDHADFVFNSYFQNHKNKGG NCYFNGAAIITEKDPSHDNCHFEYLPSN
266	sp_116800_mxdp	0.935033	0.923003	#N/A	uncharacterized protein	MLFHTKGSKNKQQTIGEKKVNKEKVALETRVFKD NSFTRGGEVKEAGEVLIYLHDEYKGRAPTINCKIE AFGKIMQRFDNRRRNWVNEMGFGGLYHLCGKH LPRNFVYWLMTRVDPVNQIFSAPNGIDFPMSKN QIRWIFGIPIKGKVIPKTENDADEEMKAKAQFILGA YGKTWESDHPNNIGAVISTDAIPVNPKFIQRLEGT WEDKDEEEFKTMFLIAAVEMVLCPTQCGRLSPS MLYACSLGMQAREYDWCNLVYDFFIERVKVFCR DFHTFGWTKGVGGCSMYLVIFYLDRLNRMPLQ WGVFPRLKALDINDLLKKARQADLLPSGDYGKIGV CFSYPLTISANM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
267	sp_117040_fhpe	0.9939491	0.9955558	#N/A	Zinc finger (C3HC4-type RING finger) family protein; FUNCTIONS IN: ubiquitin-protein ligase activity, zinc ion binding; INVOLVED IN: N-terminal protein myristoylation; LOCATED IN: plasma membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841), Zinc finger, C3HC4 RING-type (InterPro:IPR018957), von Willebrand factor, type A (InterPro:IPR002035); BEST Arabidopsis thaliana protein match is: Zinc finger (C3HC4-type RING finger) family protein (TAIR:AT2G38970.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MRRRWRQRLINRNASDFAVETTNNNDAHREFTV KYFISDGKCSICSESLKEGEDFALYISECSHAFHY NCILYNALHGNNICPTCNSLWKNLPFLNASPTPY PNNKIHIPNPPPPPNFSPKFPPIRHLPYADDDPLS SSTASSSFPNNNNHIPKPPPPPPKFPPIIRPLPFAD DDPLPSSPATTTASSSFPQLSESDTTLYPTNNVHI HIPKPPPPPPPKFPPIIRRISFADDNPLPSSPVTTIA SSSFPEHVALRESDTTLYPTNNIHNHIPRPPPPPT FPPIHSYQVLSSTFTTTTTATTTASSSATPLQQQ GANMTVKAFPEYPALGKSEQSNSFAVLVGIQAPP LSPEASDRSTDRAPIDLIAVLDVSGSMTGNKITLL KKAMVFVINNLGPTDRLSIITFSTKARRVTPLWRM TDTRKNDSIHAVNSIVVEEATNIIAGLKMAVQVLD QRREKNPVSSILLLSDGQDNENEPGKHFLRLLPY CIRSNDTDSSNSAHHQTDAVIPVHTFGFGSDHDA TTMHAIADISRGTFSYIESIRIIQDAFAQCIGGLLSV VAQQLEIQVETASVKVRIQCIPSGKHRNNIGYSKE HGVIYLDNVYAEEVKQFLVYLSVPPAELGTLTTQL LEVKCLYMDPLSKQNKVSETVKVEIYRPDVEDLS ETDRQVCLEVSKEKHRITVAKGIAQAQAMAERG QLHEAQSSLQNQITFLETVRDSGFDAAHCEMLTE EVKVVKQMMASKGVYEGTGRGFTKSMMLCHGY QRAATQQSSKAPAGRSVAMSYQTNNMKKMVSK SRRMRQGDDDQDSIP
268	sp_117750_hpak	0.932267	0.9311037	#N/A	NB-ARC domain-containing disease resistance protein; FUNCTIONS IN: protein binding; INVOLVED IN: apoptosis, defense response; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing protein (InterPro:IPR015766), NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: LRR and NB-ARC domains-containing disease resistance protein (TAIR:AT3G14460.1); Has 25652 Blast hits to 19792 proteins in 881 species: Archae - 38; Bacteria - 1467; Metazoa - 4997; Fungi - 304; Plants - 18213; Viruses - 0; Other Eukaryotes - 633 (source: NCBI BLink).	MAEAILPMAKLILGKIASGPSDLVYTYVTEEIKGA KSAKKDLKIIADKMEAICAVLEDEEEEFSNKANKI WLRDLKSIVYDIDDLLDEVAFDALQRRVNQGYFG RQLRYYLSSSNPLISRFDLSHKIESLRKKLEIIVAK KNEFDLTKRPVKASKVEIRDPFDECDVYVEPAIIG RNEAKQDIISKLTALGDASELSVFSIVGMGGIGKT ALAKLVHHDVPNFDLKLWVCVSDKFHIPKIIEDILK AGTGDNTPNQDMSMLVKKLGNLLNGKKYFLVLD DVWVEEIDEWRKLKNLLAVGKKSGSSILTTTRLAN VALMAETMESYNLDCLSDHVCWSIFKQLAFKEG EDGRYPNLHAIGWCIVKKCRGIPLVVKCLGNLLR KLTSLPNSIKYLQKLQIKNCEELDFDKGEGLEGLQ SLRLLYISNLKLTRLPNDIQSAAASLQHLSVENCL GLVELPYWLHRFASLRTIEIIYCRNLLSLPKTISHLI SLQELKIKGCPHLSKRCGVPGGEDYSLVQHVPK FELDEVCFLTK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
269	sp_142090_cuqd	0.936305	0.933403	#N/A	PEANUT 1 (PNT1); FUNCTIONS IN: mannosyltransferase activity, transferase activity, transferring glycosyl groups, alpha-1,4-mannosyltransferase activity; INVOLVED IN: GPI anchor biosynthetic process, cellulose biosynthetic process, plant-type cell wall biogenesis, embryo development, cell division; LOCATED IN: integral to membrane, endoplasmic reticulum membrane; CONTAINS InterPro DOMAIN/s: Mannosyltransferase, DXD (InterPro:IPR007704); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVINFGLVLIFSTILRVLLILYGDWQDTHMEVRYT DVDYLVFSDAANLIFSGESPFKRTTYRYSPLLAFL LIPNAILHRSWGKYLFSAADLLVGYLIRTILKLRGV PEKICTYCVMAWLFNPFTFTIGTRGNCEPIVCAVI LWVLTCIMRGKVYQAAVWYGLIVHFRIYPIIYALPI VLVLDRQYIKSGQKPMLVKWSSHQGNSNQCLG DTYPKSSWSVLASIFSGTRILFGLMSGAVFLGFT GLFYYLYGWEYLHEALLYHLTRTDPRHNFSIYFY HIYLHYEQELSILEKLASFLPQFAVQLVLMFNYAL DLPFCFLVQTVAFVAFNKVITAQYFVWFFCLIPIVL PWSKMKLKWEGVICILVWMGIQLNWLMWGYLL EFKGKNVLLPLWGASLLFMAANCYVLIVIIRCHRY SSIFQHLNSSSKERLKVK
270	sp_176320_kfpu	0.930084	0.971955	#N/A	Tetratricopeptide repeat (TPR)-like superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT4G21065.1); Has 36578 Blast hits to 13720 proteins in 224 species: Archae - 0; Bacteria - 18; Metazoa - 87; Fungi - 88; Plants - 35844; Viruses - 0; Other Eukaryotes - 541 (source: NCBI BLink).	MSKEGVPPDRYTLPITIKAVGQAYSLKTSQQLHG VAICHGVEDDEFVESGLINMYVKAGEFRDAYKVF DGSPVRNLGSWNAIIAGLAQSERYRDAVDVFLR MRRSGFEPDGVTMISVVSACGGLGDLRLALQLH KCVFQVKWKGEEKVMMMNSVIDMYGKCGRMDL ANCVFLEMNEQNVSSWTSMIVGYATHGLAKEAM EFFHAMIEACVVPNHVTFVGVLSACVHGGKVDE GKRCFQMMRNLYGIEPNVKHYGCMVDLLGRAGL LEEARNMVETMPMKPNGAIWGCLLAASERYNDV KMGEWVAKLFLELEPWNDGVYVVLSNIYASRDM WEEVEKMREIMKQRRLGKLQGYSSVTLSIIN
271	sp_177950_zogd	0.954098	0.937562	#N/A	sodium/calcium exchanger family protein / calcium-binding EF hand family protein; FUNCTIONS IN: calcium ion binding; INVOLVED IN: transmembrane transport; LOCATED IN: vacuolar membrane, plasma membrane, vacuole, membrane, plant-type vacuole; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: EF-Hand 1, calcium-binding site (InterPro:IPR018247), Sodium/calcium exchanger membrane region (InterPro:IPR004837), EF-HAND 2 (InterPro:IPR018249), EF-hand-like domain (InterPro:IPR011992), Calcium-binding EF-hand (InterPro:IPR002048); BEST Arabidopsis thaliana protein match is: Calcium-binding EF-hand family protein (TAIR:AT1G29020.1); Has 622 Blast hits to 603 proteins in 166 species: Archae - 12; Bacteria - 69; Metazoa - 8; Fungi - 194; Plants - 285; Viruses - 0; Other Eukaryotes - 54 (source: NCBI BLink).	MGMLAGSTIMLLTVIWGCCVFVGKCDIQNSRALN GIDTKGFHLQDSGVSTDIWTSYAARIMAISVVPFL VVQLPNALNSTSGRHYSILISLIVSLLLLIGYCVYQ VSQPWIQERRLDYVKHKRVMTGFLKHLETHATG KFYNRDGTINRDVVEKLFKAIDQNGDNHLSSGEL RAFVVGMQFEGMGEDYFLDKVMKEFDTDRPDD RIDQQEFVQGISKFLALVRGEKAIDHNDADTQKYL SHYDEQATKEHFLVDDFLDEEKGEEIENRKATTIK AILLILLGTVITVAFADPLVDSVHSFSEATNIPPFFI SFVFLPLITNSKGAVSVISFASSKKRKSVSLAFCE LYGEVTMGNVVGLTVLLTLVYARGLTWDFSSEVL IVFLVCTLMGALTCFRTTFPLWTALTAILLCPFSLL LVYVLDNFFGWS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
272	sp_187200_zfmf	0.997736	0.922528	#N/A	plantacyanin (ARPN); FUNCTIONS IN: electron carrier activity, copper ion binding; INVOLVED IN: pollination, anther development; LOCATED IN: middle lamella-containing extracellular matrix, apoplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Plastocyanin-like (InterPro:IPR003245), Cupredoxin (InterPro:IPR008972); BEST Arabidopsis thaliana protein match is: Cupredoxin superfamily protein (TAIR:AT5G26330.1); Has 1174 Blast hits to 1133 proteins in 56 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1174; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGQGRGSARSALVVGLAILCLLAIVQPTLAAVYNI GWSFNVNGARGKSFRAGDVLVFKYIKGQHNVVA VNGRGYASCSAPRGARTYSSGQDRIKLTRGQNY FICSFPGHCGGGGMKIAINAK
273	sp_198600_jqjw	0.92785	0.958639	#N/A	cyclin-dependent kinase B2;2 (CDKB2;2); CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: cyclin-dependent kinase B2;1 (TAIR:AT1G76540.1); Has 109034 Blast hits to 107606 proteins in 3051 species: Archae - 93; Bacteria - 11427; Metazoa - 41458; Fungi - 12141; Plants - 24398; Viruses - 437; Other Eukaryotes - 19080 (source: NCBI BLink).	MEATATGGEKIALSAMDAFEKLEKVGEGTYGKV YRAREKATGKIVALKKTRLHEDDEGVPPTTLREV SLLRMLSRDPHVVRLLDVKQGQNKEGKTVLYLV FEYMDTDLKKYIRSFRNTGDTVPPPIIKSLMYQLC KGVAFCHGHGVIHRDLKPHNLLMDRKTTILKIADL GLARAFVIPIKKYTHEILTLWYRAPEVLLGATHYS CGVDMWSVGCIFAELLNNQALFCGDSELQQLLHI FRLLGTPNEQIWPGVTKLKDWHEYPQWSPKDIS SAVPGLDKEGLDLLSKMLIYEPSKRISAKKAMEH PYFDDLDKTHL
274	sp_206500_rysz	0.976179	0.935807	#N/A	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT1G49390.1); Has 8498 Blast hits to 8450 proteins in 980 species: Archae - 0; Bacteria - 1051; Metazoa - 121; Fungi - 1005; Plants - 4926; Viruses - 0; Other Eukaryotes - 1395 (source: NCBI BLink).	MAASSVSPMKPELPPKPTQELSKITGDQLPDRYI HQFPETAADHSAIKYMDNPIIDLNLLSAVSTHQPE ELQKLRSSLSSWGCFQLRNHGLTSSLLHQIRQV EKEFFGLPLEVKQKHSRTLHWVEGYGNDTVSED QKYCWNDRLVLKVYPVDQRNFKFWPEDLPKFR ETLDEYTKEVRRVLEMILKAISKSLNLDENSFVRE CEGDEGIYMLARFNLYPPCSTPDHVVGLKPHSD ASAVTILLQDNEVEGLQVQKDNQWFKVPIVPDTL FINVGDQLEIMSNGILKSAVHKVVIDKERERTSVAI ACFPYIEKEIGPISELIDKERPQLYKKVKNYISLYLQ YSPTAERAITAMKL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
275	sp_215910_xzds	0.9484432	0.9661388	#N/A	Cupredoxin superfamily protein; FUNCTIONS IN: electron carrier activity, copper ion binding; CONTAINS InterPro DOMAIN/s: Plastocyanin-like (InterPro:IPR003245), Cupredoxin (InterPro:IPR008972); BEST Arabidopsis thaliana protein match is: Cupredoxin superfamily protein (TAIR:AT5G26330.1); Has 1434 Blast hits to 1387 proteins in 64 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1434; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGITQTSKSLYNILNLMKLILLISISLVFTIAKSDVIA VGGENRWSDGFNYVSWAQTLNFTVGDVLEFSY TRGSHNVLDVTSATFRSCDNSTGVNAMYSSGDD HVQLNEAKQYWFICTVREHCRLGMKFGINVTTTS LNRTSNSTLNDPSSGNKHINSGFYLLAVWIFSMF L
276	sp_014560_mxdg	#N/A	0.947361	#N/A	beta glucosidase 16 (BGLU16); FUNCTIONS IN: cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: carbohydrate metabolic process; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 1 (InterPro:IPR001360), Glycoside hydrolase, family 1, active site (InterPro:IPR018120), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: beta glucosidase 13 (TAIR:AT5G44640.1); Has 11475 Blast hits to 11098 proteins in 1475 species: Archae - 144; Bacteria - 7936; Metazoa - 718; Fungi - 200; Plants - 1457; Viruses - 0; Other Eukaryotes - 1020 (source: NCBI BLink).	MAVLHSAPFQALHSSVTLKNPNSSLFISRTNTSK YYLPTKISCSLIPNVIKKAAKTVVNPVISIVDGLQPA TFTPPKDDLKKEDFPPGFLFGASTSALQTEGRGD EGGRGPSTWDFMIKADKGHKAVDSYNLYKEDVK LLKETGLNTYRFSIAWPRILPKGTIESGINQEGIDF YNRFIDELIANGITPIVTLFHFDLPVSLQSEFYGFM RPNIVDHFKEFADLCFEKFGDRVKYWGTINEPYV FGSYGYKMGLPNALKDDPLGTYIATHHILAHAAA AKLYKEKYQKSQGGKIGISLPCKWYMPHASADP KDFVATKTQLDFTLGWYMDPLTRGDYPESMKKS VKGLPVFTKAEKELIKGAFDYIGINYYTSRYVEHT CVVDAMVGTITIKYSEQKVDDEGNLVGWQAKGQ EDIYVRPDGLKNLLLYVKTEYDNPEVYVTENGVP DKPTITSSIEDIIAQGNTTELEAALYDDYRIQHTED HLVAVRGAIREGANVKGYIVWSLMDNMEVGAGY ETRFGLNFTDYNDNHKRYPKKSAKWFANFLKEE TTTNGSLKKN
277	sp_017220_zcfr	#N/A	0.943837	#N/A	VPS9B; CONTAINS InterPro DOMAIN/s: Vacuolar sorting protein 9, subgroup (InterPro:IPR013995), Vacuolar sorting protein 9 (InterPro:IPR003123); BEST Arabidopsis thaliana protein match is: Vacuolar sorting protein 9 (VPS9) domain (TAIR:AT3G19770.1); Has 1393 Blast hits to 1376 proteins in 227 species: Archae - 0; Bacteria - 0; Metazoa - 508; Fungi - 217; Plants - 509; Viruses - 0; Other Eukaryotes - 159 (source: NCBI BLink).	MSNLKFQFTQFYISQTPQNHTSNQTHAHFPIIPHL NSPPFKRRTRTITCGLRKIGPRKPLWRSSFLSTE AIQAVQSLKLAKSPVKLQQVFNTRITRLLKADLLG ALSELQRQNHLDLALKVFDFVKKEYTADLSLYYD MILLLGKNKLIQKAEELFDELKKEGLQPDVRAYTE LIGAYFQVDMVEKAMETYNEMKASGFTPDKLTL MIMVRNLEKAGEQELAASVKRDFEEYVDEPERFL CQVEKNYSYSTDL
278	sp_023330_reqs	#N/A	0.953864	#N/A	uncharacterized protein	MILVGDEEDGEDLKGSRLRHMRLTTDQKDEEDG WALASVFGFSGVEEIYVARRYEQNQRVTEFQIEI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
279	sp_031030_kwtl	#N/A	0.945223	#N/A	embryo defective 2739 (emb2739); CONTAINS InterPro DOMAIN/s: Integrator complex, subunit 3 (InterPro:IPR019333); Has 179 Blast hits to 170 proteins in 63 species: Archae - 0; Bacteria - 0; Metazoa - 131; Fungi - 0; Plants - 37; Viruses - 0; Other Eukaryotes - 11 (source: NCBI BLink).	MKYLHAIVNDGYCLFIHLISQIVDGLYSKMLEFPKV QLIWVASVMIDVSGVGFDSLLVSLLRQIVGGDFS EGNLWLSFEIVNLFTSKWDSILDEEPLVLSSALYT YLRLLADHYRLAQSPKLIVLKRKEIGFCVRMLREK FHLCLRIGRDLVRLLQDLVHIPEFRAIWKDLLLNP NSFQTSGFTDISHIYGMKTSSRYFLLRITPEMETQ LRFLLTHVKLGSQRRHQAWFTRKFLFAPERETLV SDIVRFICCSHHPSNEILQSDIMPRWAVIGWLLKL CSKKYFEANVKLTLFYDWLFFDEKVDNVMNVEP AMLLMVNSIPRYIDITNSLLEFLLLVVENYDMDRK NVIYNGVSSAISMLVSKGVIGSLDIFSSCKLLSPFV KEMLQRFMSKVQSGNPSPEMPPEYLSSEMPPQ TSPNPPYLEIQTHQSAKLRACPRNAELSAQHISVL KPTSAASLCEGQVDNLTNLIRKIREFMIISSKEGQ KLVENILRLYINPANEGKFGGIGSEYLACELKKEL ELAGYNLFSPVGSLPNSDDNADDEVQSVTAAIIR TFIISQHQSIRLMLFWWSRNGCLVGPRLLSYASR LAHEAHEAGYGASYTTDINYTKMPLLKNHIEGYM SLTTSGNDDHVNGTITPNPEVDKKLISKMIDSAFA SYRSIFSHAKDSLSKEVGIVPGKLLSMDVISCSLW KSKFFKDLFCYIFSNLPDISLSDEDVILLLVTQLED TDLLYMQMDLGLKRVTIFGEDVKKVLQLIRRSIYW KQVDQHNVWGLIRSELTVSEFPVEKLLLGCFFST DFDSKTSGIVVGGHLALCSSCAPTPEIVGAVMLL PNDKFPGFAGAIFSHWTVTNASMFCNSVADFLN KLGNKNGKSPLDWDGFLINQSGLLFLLEYFKAQG ERGVNMLEKLALILSNLEIRPGETTVPMDTG
280	sp_034290_khnd	#N/A	0.935035	#N/A	ROOT UV-B SENSITIVE 3 (RUS3); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF647 (InterPro:IPR006968); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF647 (TAIR:AT3G45890.1); Has 450 Blast hits to 446 proteins in 132 species: Archae - 0; Bacteria - 0; Metazoa - 113; Fungi - 75; Plants - 194; Viruses - 0; Other Eukaryotes - 68 (source: NCBI BLink).	MAEQILVEEWAGSSSSKLSITATISSTPNISIRRFD SRFKLVRRQILSAFVPEGFPSSVTPDYVPFQIWD SLQGLSTYIRMMLSTQALLSAIGVGEKSATVIGAT FQWFLRDLTGMLGGILFTFYQGSNLDSNAKMWR LVADLMNDLGMLMDLISPLFPSAFVFILCLGSLSR SFTGVASGATRAALTQHFALQDNAADISAKEGSQ ETVATMAGMALGMFLAHVTRGHSMAIWFSFLCL TMVHIYANYKAVRCLSLKALNGERSSLLLQHFMK TGQVLSPEKVSMEEHVLPLWLNPWSSKKFKCLH SRINLGVKVSSLNHAAMTDLFFCAGSYYKKAKYL LLEKNGMISIIHKDSTPVDILQSYIHALVMANLSNK SKSGHVESFSWMDKCYEDFIKKSSGWKTERLLS HSVVWRASWFIGQSDEKID

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
281	sp_038050_rktw	#N/A	0.948018	#N/A	ARM repeat superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-type fold (InterPro:IPR016024).	MAPESNLEASPVINTLQWLEHCHENGVEVNDAE EERFQLHADENGPTHHPSAPPDEIFDISTTVDPR YIISLIRKLLPADERQGDDLLGLSVWNNSTQRSDA DQIVDGASSPKRNSGLNNLNGDLETMESQPVGL EQPSRSAGPDDNDTGLPKKTAREEVWEECGCV LWDLAANKEHAEYMVQNLVLEVLLANLLTSDSVR ITEISLGIIGNLACHEVLMDKITSTNKLLETIVDQVF SDDPPCLCEVCRLLTLGLQGSQSVVWARALQLD HILSRVLWVAGNTLNPVLIEKSIGLFLAIMENQHV VPILLPTFMNLGLPDLLVSLLAVELGKLTSDRTSE RNVFGSDNVTYLLEHLHCLKTFSSLFTPKVQGSC TDFVMYPALDSILRAVEALSVTDDYSERISSKKEV VQMAIELIKIPEKIEISSSCVTASVLLANILADSPDM AFKLSEDFLLLQSLLDLFPFTSEEFEARNALWSIL ARLLVRAQEAEMSSSRFSQFVLMLVDRSDMLED ALLESDWEKSPKRDNLTTSEAKFDSSTAAIKCIVK ILNQWNLVKDEVSVVCTDEDSVRKSYPDDRDVQ KLLDCCLKHAS
282	sp_040250_gauu	#N/A	0.938537	#N/A	plant VAP homolog 12 (PVA12); FUNCTIONS IN: structural molecule activity; LOCATED IN: cytosol, endoplasmic reticulum, nucleus, plasma membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: PapD-like (InterPro:IPR008962), Major sperm protein (InterPro:IPR000535), Vesicle-associated membrane protein (InterPro:IPR016763); BEST Arabidopsis thaliana protein match is: vesicle associated protein (TAIR:AT3G60600.1); Has 1106 Blast hits to 1079 proteins in 208 species: Archae - 0; Bacteria - 0; Metazoa - 441; Fungi - 140; Plants - 434; Viruses - 0; Other Eukaryotes - 91 (source: NCBI BLink).	MNSGEFLSIQPQELQFPFELNKQISCTMQLTNLT DSYVAFKVKTTNPKKYSVRPNTGIVLPHSNHDLL VTMQAQKEAPADMQCRDKFLLQSVIASPETTIKD ITSDMFNKESGNLVDEWKLKVSYVSPPQPPSAV HEGSDEGSVSDNGHMDTFSAAAITHVEPEDDSR QVNATISRLLEEKNSAIQQNNKLHQEMDHMRREI SKKRATIPFMYVIIVGVIGIMLGYLVKRI
283	sp_048010_ucep	#N/A	0.933203	#N/A	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein; CONTAINS InterPro DOMAIN/s: SAM dependent carboxyl methyltransferase (InterPro:IPR005299); BEST Arabidopsis thaliana protein match is: S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (TAIR:AT3G11480.1); Has 909 Blast hits to 891 proteins in 125 species: Archae - 0; Bacteria - 69; Metazoa - 9; Fungi - 5; Plants - 730; Viruses - 0; Other Eukaryotes - 96 (source: NCBI BLink).	MEIKNYLHMNVGNGELSYSQNSSVQARVWLMN QSLLENAIQSLITSKDSGLNIADLGCGVGNVPLGL VSFLVEIMQKKSKELLKLDNNDDQVPQLQIYMND LPSNDFNSLFKEMMNLDVLKKKDDDGVPLCFLM GAPGSYYDRLFPNKSLHLVHANYSLHWLSKVPP RIYDDQRMSINKGNIYISETSPAKVGRIYRDQFEQ DFTLFLKSRSKEVIANGCMLLTLRGRPSSTNPLT WTTFEFKIFTKTLASLVSEGLIKEEKLDTFDFPCY CATKEQLESIAKKEGSFEVETSKTMVVDVAPEIE DKWERAQVITKIIRAFSESLVSSHFGEDISTLLYD KLVHFANQHLVDDQLAQNHAVTVLLRKK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
284	sp_048090_axhf	#N/A	0.923115	#N/A	sinapoylglucose 1 (SNG1); CONTAINS InterPro DOMAIN/s: Peptidase S10, serine carboxypeptidase (InterPro:IPR001563); BEST Arabidopsis thaliana protein match is: serine carboxypeptidase-like 9 (TAIR:AT2G23010.2); Has 2871 Blast hits to 2808 proteins in 264 species: Archae - 0; Bacteria - 35; Metazoa - 597; Fungi - 474; Plants - 1434; Viruses - 0; Other Eukaryotes - 331 (source: NCBI BLink).	ESIYYLLENWANDIQTQAALHIQKGTKSSWIRCN MTLAYAYNVESSFGYHQNFTQDFIRALIY
285	sp_062340_nrhw	#N/A	0.94717	#N/A	DNA binding; FUNCTIONS IN: DNA binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Zinc finger, BED-type predicted (InterPro:IPR003656), Protein of unknown function DUF659 (InterPro:IPR007021); BEST Arabidopsis thaliana protein match is: hAT transposon superfamily (TAIR:AT3G22220.2); Has 518 Blast hits to 485 proteins in 17 species: Archae - 0; Bacteria - 0; Metazoa - 8; Fungi - 0; Plants - 505; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MSLGNFDLLVSCPGSLDPKDPARKYGRPDPNNR NNFTCNFCGKVTKGGAYRLKQHLVGGFRNVTKC YSCPEHVKEEVKFFMIKKQQAKVGGQMLPRVTQ SDVDGCNDDEEDCETLNSKGGSSVKATSKRPRL KGPMDMYFTPIPEDAISGRKNGNEGRQRGSFSIY VPSRSVQSGADPIASDTPGRRDLQLQGTPSHFR RNAVYVRRRRRF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
286	sp_062780_spxw	#N/A	0.955043	#N/A	Polynucleotidyl transferase, ribonuclease H-like superfamily protein; FUNCTIONS IN: ribonuclease activity, nucleic acid binding; INVOLVED IN: RNA modification; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Ribonuclease CAF1 (InterPro:IPR006941), Polynucleotidyl transferase, ribonuclease H fold (InterPro:IPR012337); BEST Arabidopsis thaliana protein match is: Polynucleotidyl transferase, ribonuclease H-like superfamily protein (TAIR:AT1G55870.1); Has 339 Blast hits to 335 proteins in 120 species: Archae - 0; Bacteria - 10; Metazoa - 154; Fungi - 41; Plants - 69; Viruses - 0; Other Eukaryotes - 65 (source: NCBI BLink).	MQKRLFCTQAAIQIENTIKNTSLSKKWIIKQVKKS NFSQVLDEIKPHISNSDFIAISLQKTGSYSSPWQR VLPFDTPEISYLKAKRSAERFQVLQFAVCPFTLRA SKLTAYPYNFHLFPRDELKIGMPSYSFACQTSYL TAMARLGFDFNSCIYDGISYISRAQESSMKHLPG NTMLRAHASQSSPAVSVADSIYIQRIKTRVKTWIK ACKDSKGLNEDPMVRALRKVIMGEHGSRPCLDL DVCSERQVQLVLEVMKEYSDDTVPLLIPAKRGGT QGLRIVLTSSEQDKILFQKELQDLEEEQSKNVRG FREVIDLISASGKPVIAHNSLDDFTFIHSKFLSTLP PTIDEFKHSLSSFFPYIIDMGHVTKEFNRSRNMNN IPAASSFLRSRFFAPIDMDIPIKGTPDEMSEGKING QNVLQLSHLFAKLNTIINRGQHHVISEDKILALDNF ANVFDPYSSPLPENSDDDDVRISSGSLRKLSCDN MVFIWGFRSGLSTGELKTLLSGSHEAFSAEFDVQ LVDKSCAVAALQQPGLAGAFVKAMESGGGCHE GGLKELLSEGLRAADYGTYKLACNMSKWDSNLA DSLDQVLMSADQFPVDNSHGRSPYEICSFNQRK QLFSISPPTRKIRGRNGGNLGMGAVCATHNFDVV VVGAGIIGLIIARDLLLRSNLSVAVVDAAVPCSGAT GAGQGYLWMVHKTPGTEVWDLAMRSRKLWQL LADNLQDQGSDPLEALGWKKTGSLLIGKTEAQL NQLKSRVQQLSMAGIRAEYLSNNDLAIREPELAIN QQSGAAFLPDDCQLDAHRTMAFIEKDNRLFLSQ GRYAEFYHDPAVCFIRSNGKGKIDGIKTQKNTLY GRKATIIAAGCWTGSLMDKLLQNSDIVLHVPIQPR KGFLLALENFNNLRVTHGLMEAGYVDHQKSVISA SDKVDSNEALSISMTVTRDASGNLLIGSSRQFAG FNTELDTFIVDRIWERAQDFFPSLKEFSVKNLSE QMNVRIGLRPYMPDGKPVIGPVPGITNLFIASGHE GGGLSMALGTAEMVTDMILDNPLQLDCKPFAVE GRCC
287	sp_063090_jhxo	#N/A	0.942262	#N/A	uncharacterized protein	MENLIRSVIQSTQRIQVSVEPDIQDATMEDVLNNC MENNALAQEREEEESTNEIEAEEQHIIYDSEEPD QDIVGTLIGYTAETVEELLVRINEKHATSLNLSNTE DGSTIETSATAAPETSTTRAPETNSAPTGTHATT TSDTSVQQAATSSEPATNTATEPPMVLDPERCT TKGRNKRPRGPFVKKKKGKTAAPPTADFGTITPN LRLF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
288	sp_065250_adcs	#N/A	0.9823747	#N/A	EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cactin protein, cactus-binding domain, C-terminal (InterPro:IPR019134), Cactin, central region (InterPro:IPR018816); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G36815.2); Has 11711 Blast hits to 7382 proteins in 452 species: Archae - 31; Bacteria - 352; Metazoa - 6006; Fungi - 1138; Plants - 599; Viruses - 33; Other Eukaryotes - 3552 (source: NCBI BLink).	MARHERSKSSSRRREKPSSSRKKQRRHSDNDSA VSQSETDYDSNSSRDPSRRRRRPSNSHRRRESS PETSEDSGDRRKRKKKVSEEEIKEYLAKKAQKKA MRVAKKLKTHTVSGYSNESNPFGDANLNEKFVW RKKIERDVCDGVPMEEFTVKAEKKRQKDRMSEIE KVKKRREERALEKAQREEEMILLQRERSRAEAD DFEKKEEEHFDQSKVRSRIRLREGRTKPIDVLT KYLDDSDEFDVELDEPYMVFKGLTVKEIEELHDDI KLHLGLDRATPTHVEYWEALMVVCEWELGESR RKDAIDRARVRGEQPPAELLAEERGMHSSIETDV RNFLEGKSYKDLVELQAQIESEMRSGKAKVVEY WEAVVKRLHIYKAKACLKEIHAKLLRKHLHRLEQ PLQDENVKSEEEKIEEEESEEEVKVCLDARAFSP EPIITHEEAQESEAEDGSFSPQLLHSEDTEEAIDP EEDRVILERKRLAVIEEQRAKEAMASKPTRAQPED DLEVRARKVMGVVEDGDAVFGAGAEINLDSQVY WWHDKYRPRKPKYFNRVHTGYEWNKYNQTHY DHDNPPPKIVQGYKFNIFYPDLVDKIKAPTYFIEKD GDSSETCIIRFHAGPPYEDIAFRIVNKEWEYSHKK GYKCTFERGILHVYFNFKRYRYRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MSRKNDGPSSSINKPLLENHQPKPNVKINLGPKL
						TLISSGTLAVLYVVLSILTFVGNDNKGFSSSSWEK
						AEVMFRIVQALTQVAIFVLVVHEQRFHATCHPMS
						LRFVWIAQFVVVLVFSISGLTRMLSAGEELDSFM
						RTDDMFSLVSLPFCFYLLIVASQTSSGIVVIKEEN
						GLSGRTEVNETIVDDPNVSDYAKANLFSKTMWL
						WMNPVINKGYKNPLKIEDVPSLPPDHLAERMSKL
						FESQWPKSGEHTKHPVAVTLLKCFWKNIVYTGC
						LAVIKLGVMYVGPVLIQSFVNFTSGMGSSPYEGY
						YLVITLLIAKAAEVLSAHHFNFHTQRVGLLIRSTLIT
					multidrug resistance-associated protein 10 (MRP10);	SLYRKGLKLSCSSRQAHGIGQIVNYMAVDAQQL
					FUNCTIONS IN: ATPase activity, coupled to	GDLMNQLHSIWLMPLQLFAALALLYFNMGVSAIA
					transmembrane movement of substances;	TFIGVIVILGFVFIRTKNCNVFQYQVMRNRDLRLK
					INVOLVED IN: transport, transmembrane transport;	ATNEMINNMRVIKFQAWENYFHKRIQSYRELEYG
					LOCATED IN: vacuolar membrane, plasma	WLSKFSYTVAANMGVLWSTPVALSVLAFGVAIW
					membrane, vacuole, plant-type vacuole;	MGMYLDAGTVFTVTTIMKILQEPVRNFPQSLIQIS
					EXPRESSED IN: 25 plant structures; EXPRESSED	QAMISLGRLLDGYLMSLELDQGLEEREEGCQDDIA
					DURING: 13 growth stages; CONTAINS InterPro	VEVNNGSFSWEDEATEQTLKGLNFQVKKGELSA
					DOMAIN/s: ATPase, AAA+ type, core	IVGTVGSGKSSLLAAILGEMHKISGKVKVCGTTAY
					(InterPro:IPR003593), ABC transporter-like	VAQTSWIQNATIQENILFGLPMNKEKYKETIRVCS
					(InterPro:IPR003439), ABC transporter,	LEKDLEIMEFGDKTEIGERGINLSGGQKQRIQLAR
					transmembrane domain, type 1 (InterPro:IPR011527),	AVYQDCDVYLLDDIFSAVDAHTGSDIFKECVRGA
					ABC transporter integral membrane type 1	LKDKTILLVTHQVDFLHNVDQIFVMRDGMIVQSG
					(InterPro:IPR017940), ABC transporter,	KYNEILEAGLDFGALVAAYENSMELVETSSHATE
					transmembrane domain (InterPro:IPR001140), ABC	DDTPKSPISENKTIERVKSEKKALTKTDSQKGTSK
					transporter, conserved site (InterPro:IPR017871);	LIEDEERETGRVSLEVYKKYSTEAFGWWGVSAV
					BEST Arabidopsis thaliana protein match is: multidrug	VIVSLFWIVSLLGSDYWLAYETSEERTFTPSLFIIV
					resistance-associated protein 4 (TAIR:AT2G47800.1);	YAIVGIISCLLVTARAFVITFLGLKTAQGFFNQILDS
					Has 708251 Blast hits to 364353 proteins in 4024	ILHAPMSFFDTTPSGRILNRASTDQTNIDLTIPLFL
					species: Archae - 1314; Bacteria - 571430; Metazoa -	GMTIVMYFNLVSILVVTCVYAWPTVVLLVPSIWLN
					13503; Fungi - 8752; Plants - 6721; Viruses - 16;	LWYRGYYIAISRELTRLGSITKSPVIHHFSETVSG
289	sp_068740_kuop	#N/A	0.932489	#N/A	Other Eukaryotes - 94685 (source: NCBI BLink).	VMTIRCFRRQALFCKQNIEKVDTSLKMDFYTNGA
						MAEENNTSESHEKGHQHLELPEYDIGRQSTTYKL
						TLSMKRRFSPQVVGAQTPSPEGAGEVRDLLRRL
						VEIDNESRKLLVKVFSLLEKDESLESYVKKRIEDP
						KVSRMWLLVEQVLACQEGDTNFNQVIMGDDNSI
290	sp_074060_fdfx	#N/A	0.922246	#N/A	uncharacterized protein	EDGNYTTYDVKSIKEKDEIGEEKRNERFKVRNITS
						GHQPVPGVTTITANGK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
291	sp_079500_yqfd	#N/A	0.928585	#N/A	Exostosin family protein; FUNCTIONS IN: catalytic activity; LOCATED IN: membrane; EXPRESSED IN: stem, root, stamen; EXPRESSED DURING: 4 anthesis; CONTAINS InterPro DOMAIN/s: Exostosin-like (InterPro:IPR004263); BEST Arabidopsis thaliana protein match is: Exostosin family protein (TAIR:AT4G32790.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGFELLKAFQVETRRLLWLIGLVFITILFQCIELPK GNSLWSLFSRGDEIPEKKVLSFFGDNSTSYSASL SNMSSPSTSIVSEDRNNSSVRIHLCNTTECGSNT SSNEAPKKDFNLAPSEPPVAPFLLSHDVASNLST EFSAPPLPLIHVSENKTYISVNPELGNTTERSSAF NEAPKEDFALPPGHNGDSNLAPVESHDVGRPNL SMEFSTPPPPPPPPPPPPPPPIHVSAKKPKSKMF DMKVLTLSDMSEELLQNYLSPPTMVPRWSSKVD EELVLAKQLISNAPPLKDVLYAPLYRNASMFHRS YELMEQMLKVYIYKEGDRRIFHRPPLEGIYASEG WFMKHMEANRHFVTKRPKEAHLFYLPFSSRQLE ETLYVPNSHSHKNLIQHLEDYVDMISSKYAFWNR TEGTDHFLVACHDWAPAETKRQMGNCIRALCNA DVKEGFKFGKDVSLPETYVRKARNPLRDIGGKPP SKRKTLAFYAGNMHGYLRPILLQYWGNNIDPDM QIYGSLPRSKDNRNYINYMKSSKYCICPRGYEVN SPRVVEAIFFECVPVIISDNFVPPFFETLNWESFA VFVAEKDIPNLKNILVSIPEKTFFRMHMRVKKPGL CLAQENMGFKLLKAFHVETRRLLWLTGLVFITVV LFQCIELPKGNSLWSMFSPGEVSEQKVNFFGKN STSYLVSNMSSPSAVNSKIVSEDGNNFSGRVDE GRNNVPFISENPELGNTTKSNSSFTESLKEYFAL HKGNSNLALIKPLGAPLVLSHDVAPNFSMEFGTS APPAPPPLVHVSAKNPRLTVFEKEVLSLSDMSEA LIQNFASPSTMIPRWSSKADEELVLAKQLIESAPP LKDGLYAALYRNASMFQRSYELMEQMLKVYIYKE GDTRIFHRPPLLGIYASEGWFMKHMEANRHFVT KKPKEAHLFYLPFSSLQLEETLYVPKSHDRKPLIQ HLEDYADMISAKYPFWNRTEGTDHFLVACHDWV PSETKRQMANCIRAICNSDVKEGFKFGKDVSLPE TNVRKSQNSLRDVGGKPPSKRKTLAFYAGNTLH GYLRPILLQYWGNNTDPDMQIHGTLPRSKDNRN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
292	sp_082830_szmt	#N/A	0.8311214	#N/A	DIE2/ALG10 family; FUNCTIONS IN: transferase activity, transferring hexosyl groups; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system, integral to membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Alpha-1, 2 glucosyltransferase Alg10 (InterPro:IPR016900), Glycosyltransferase, ALG10 (InterPro:IPR007006); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGRTVVALIVSLWMIPVSILVNRIVPHSYMDEIFHI PQVQQYCKGNFRSWDPMITTPPGLYYLSLAHIAS LFPGMIWVQRISSFSDLCSTAILRSVNGALAIVCSI LIYEIITQLKPALDRRKAALNTVVLSLYPLHWFFTY LYYTDVASLTVFLAMYLTSLKKNYFLSALMGALAV CIRQTNIIWMLFVACDGVIDMLLVNRQSVGIQKSD ISIAEKNRISPKDNFSAKPGVRRRKMKGTQDVPF PTVTGGNEATLSGSHLSGLFDEFQYIIIQAWHLK WELMVSFTPFILVLGAFVSFVYWNGSVVLGAKE AHAVSPHFAQTMYFAMVSALFSAPFHFTLGHAT ALLRSFFKIRPLSFFLGFVILTAGVLSVHFFSIAHP YLLADNRHYPFYIWRKVINAHQLMKYFLVPFVYY SWLSILHILGKARSRIWVLAYFLACAAVLVPTPLV EFRYYTIPFFIFMLNSGIEEISKWLLIAILYVVVNSF TMYMFLFRSFEWSHESGVQRFIW
293	sp_087510_gkqg	#N/A	0.8283353	#N/A	respiratory burst oxidase protein F (RBOH F); FUNCTIONS IN: NAD(P)H oxidase activity; INVOLVED IN: in 10 processes; LOCATED IN: plasma membrane, integral to plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ferredoxin reductase-type FAD-binding domain (InterPro:IPR017927), Cytochrome b245, heavy chain (InterPro:IPR000778), EF-Hand 1, calcium-binding site (InterPro:IPR018247), EF-hand-like domain (InterPro:IPR011992), Ferric reductase-like transmembrane component, N-terminal (InterPro:IPR013130), NADPH oxidase Respiratory burst (InterPro:IPR013623), Ferric reductase, NAD binding (InterPro:IPR013121), EF-HAND 2 (InterPro:IPR018249), FAD-binding 8 (InterPro:IPR013112), Riboflavin synthase-like beta-barrel (InterPro:IPR017938); BEST Arabidopsis thaliana protein match is: Riboflavin synthase-like superfamily protein (TAIR:AT4G11230.1); Has 2419 Blast hits to 2261 proteins in 348 species: Archae - 6; Bacteria - 265; Metazoa - 742; Fungi - 676; Plants - 525; Viruses - 0; Other Eukaryotes - 205 (source: NCBI BLink).	MRGYARHELRRWGSDNIRTPAMAMSTGSSPGS DFTTSGYSTPRSTATAEEFVEVTLDIEDDETIILRS VEPAATPPPPSTMNAESSSSFRAPNSLASMAGME IPTASTSRTPAIRRTTSSSSRIRQFSQELKAEISKA KQSFSAQLTKRFSGKFLNYSPSMKGSDAVVDPAI AARNRRKQLAQLNRTRSGAHRALKGLRFISNKG SASCSEVESNFEKLAKDGFLYRTDFAKCIGMEKS EEFALEMFDALSRRRRLKVEKISKDELYEYWSQI TDQSFDSRLQIFFDMVDKNEDGQLTEEEVKEIIML SASANKLSRLKEQAEEYAALIMEELDPEGLGYIEL WQLETLLLQKDTYLSYSQALSYTSQALSQNLQGL RYKTPMQKMRTNMFYFVQDNWKRIWVLSVWILL MTGLFIWKFLQYRNKAAFQVMGYCLTTAKGAAE TLKLNMALILFPVCRNTTWLRSTRLAYTVPFDDN INFHKTIAASIVLGIILHVGNHLACDFVRLERSSPQ DYALYLIDFGKTKPTYGDLVRGVEGITGILMVVLM AIAFTLATRYFRRGDIKLPKPFDRLTGYNAFWYS HHLLIIVYICLIIHGIQLYLVHKWYKKTTWMYLAVP VLLYAGERTLRFFRSGLYPVRLLKVAIYPGNVLTL QMTKPPQFRYKSGQYMFVQCPAVSPFEWHPFSI TSAPGDDFLSIHIRKLGDWTEELQRIFKAACETPI PGRSGLLRADETTKKSLPKLLIDGPYGAPAQDYR KYDVLLLVGLGIGATPFISILKDLLNNTVKKEEDEE KAADLYPVPDLSEGSPKQKKPFRTTNAYFYWVT REQGSFDWFKGVMNEVAELDQRGVIEMHNYLT SVYEEGDARSALITMVAALNLNHAKNGVDIVSGTKV RTHFAKPNWKKVLSKIGSKHANARIGVFYCGAPV LAKELKVLCNEYNQRGTTKFEFHKEHF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
294	sp_093090_ycwu	#N/A	0.931235	#N/A	AAR2 protein family; CONTAINS InterPro DOMAIN/s: AAR2 (InterPro:IPR007946); Has 319 Blast hits to 307 proteins in 161 species: Archae - 0; Bacteria - 3; Metazoa - 130; Fungi - 91; Plants - 45; Viruses - 0; Other Eukaryotes - 50 (source: NCBI BLink).	MEPEKALELVKQGSTLLFLDVPQFTLIGIDTQVYS AGPEFKGIKMIPPGVHFVFYSSSTKDGKEFSPIIG FFIETSTSEVVVRKWNLEEERLVKVPSEEEERYA QAVKQFEFDRHLGPYTLSDYGDWKRLSNYITKD VMERIEPIGGEITVACESSLVGNIPKTTMEVALAE QLKNSKFHREAESSENKGCYYTPVPWVIKQKGIS GAELTSVNLDKTHILETILMKECGGNEDRFLGEL QFAFIAFLMGQSLDGFLQWKSLVSLLFGCTQAPF HTRSRLFVKFVKCVYYQLKYGFQRDHARSCGIEK ESSLLLDDSWLTSDSFLHHLCKEFFLLVDEASVL DGDLLTWTRKLRELLENTLGWEFQRKSDMDRLY FEEDDEYAPVVVEMLNESN
295	sp_096790_omah	#N/A	0.924402	#N/A	uncharacterized protein	MAAAGPRPKRPQLASKNNPTRDRTTGFGGGPG GYGSKGASEEGVNHRSTRPNQQPRPPTPMRSK EALRVPIRRGGGFWFAGFGCRRYTGLWTSNQL PTATTSRGRRCCDPRGRAASNLYGEYVEIKVKS EQILQYDLLLYLKLEVVDGKLNWRSSHSDEEEEK SRVITKVKKVVKHAKEKKVVKGASKPPGARKS MRIATNEPKLIHSPTYVEISCDDEDEEGEQGEEDI EIKDKEQASKVPQPKDVVPNPRLVHVPPPTVVPI GSSGFGGFSSFVSSSSSFGSQKDPSMAELKEKV GQLKERMARMLEIQEGTDARLGELLAQVGALVS VIQEMGSGLQQLNTQVTTRSSTILQNLEH
296	sp_103010_ortt	#N/A	0.924634	#N/A	LEAFY PETIOLE (LEP); CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT1G28160.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSKTKSTKTTSPQQQDVGNPPIKFLGVRRPWG RYAAEIRDPSTKERHWLGTFDTAEEAALAYDRAS RSMRGPRCRTNFVYSDMPHGSSVTSILSPDNQN QHQNQNQNQNHHQNQPEFSLSPGNDYDYKAVS NAQVWSSSSSSSSYQENPTIFTYNNGADDSVINTN YSNNNCFYANSNDIELPPLPDQFTGELAQLSWPS PAVGGNSSEYLGLDSNNAGYVYYADVPPFARIAY HGDLQGLVRTFASQGETPQVIDVDADPSKLTAD PPNSTQIMFGSIPATISMDTKASGSANAKDTNPA DGSKGAGTSSTNE
297	sp_110710_zzug	#N/A	0.923889	#N/A	WRKY41; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent, regulation of transcription; CONTAINS InterPro DOMAIN/s: DNA-binding WRKY (InterPro:IPR003657); BEST Arabidopsis thaliana protein match is: WRKY family transcription factor (TAIR:AT4G23810.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEEVMGFDRSAHLINELLQGMEFAKELRNRIMNP ENKGCLSDDNDDQVALNNSLLEKMLSSFDKTTI AKLLHIQLPLHPSESPHSLTISSPISENSHPDDSPP SKRRRVMMRKWTKIVKMSSERGVEDGFSWRKY GQKDILGAKFPRGYYRCTHRHTKGCAATKQVQR SDGDESIYQIMYRGEHTCFQGAQAQFDDIIPAQM KSEQQPPPSPKEQEVYVGLSPEGELKVESEQEP ENRAQYFRSLSFTPANLQVVQPDPIESYLMSYCS DFDSPTTSVSNYLQVLPYNHIGLTVETSGTEFTE AVSGPNSVTNSPMMDDFEFPIEELLDFGGNFSFE L

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
298	sp_118630_cokg	#N/A	0.455665	#N/A	nuclear factor Y, subunit B11 (NF-YB11); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: intracellular; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Transcription factor CBF/NF-Y/archaeal histone (InterPro:IPR003958), Histone-fold (InterPro:IPR009072); BEST Arabidopsis thaliana protein match is: nuclear factor Y, subunit B1 (TAIR:AT2G38880.7); Has 148542 Blast hits to 60498 proteins in 2540 species: Archae - 708; Bacteria - 15552; Metazoa - 38569; Fungi - 16830; Plants - 7103; Viruses - 1761; Other Eukaryotes - 68019 (source: NCBI BLink).	MGKEKATAVATEAEELPKTIVRRVVKEKLSQISTD GGDVTLHKEALIAFSESARIFIHYLSATAHDICNDS NRQIINAEDVLNALEEVEFSEFVEPLRAALDEFKK KAAGKRAGSAKAKEASKKRKTDEETATKNVKAE VTATENGEEEDTATENGKDEETATENGEEEDTA TENGKDEETATENGEEEDTGTENGEEEETATEN GEEEETATENGEEEDTATENGKEEENEDEEDEE EEESENNDDNDDDDDSPDS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
299	sp_120540_rwnz	#N/A	0.948152	#N/A	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT5G55840.1); Has 34111 Blast hits to 7189 proteins in 218 species: Archae - 2; Bacteria - 9; Metazoa - 117; Fungi - 126; Plants - 33142; Viruses - 0; Other Eukaryotes - 715 (source: NCBI BLink).	MLHRIASTHLLTSIVNQVCSSSFFFPLKYHQFILSA FLSNSSSSNKSSNTHFNAPTITAPLVGNLVFSKYS HFWEHRIDRYTIANSSLNQLILNISYVCPEICRKFL RVSSWEAENVLEVLLGFQLGSGNLRIDVEVVESL WEIFRRVSVRNGGFKHLPRSYEVMASVLVNVGM FREVESLLSTAETRGILLANEEISGGIVEGYINLGE LGKAILMYDTMRELCLAPSLSCYRSLLGLLMKED KINQAIRVYEDMVDVGFEIGDVDAKNLGSLIGFLC WEGEIREARKFLKKIVGSGWKPSTEILSEIADVYY EKKDFEDLLSFFIEMKCAPDVLVGNRIISSLCSNF GIGRADKFMMELENLGFVPNEITFGIYICWSCLD GRIRDAFCYLSEILSRGLKPDRHSYNALISTMFKE GIQDHAREILNEMIDTGIKPNLATYRVLLAGYSMS RQFNEVQLMMREMVNHGLIVLSPTDNPLSKAFAI LGLDPLTVKVKRDNDVKNFRTEFVDALGNGLYLD TDVDNFEKTLTGVLQESLIPDFNLLVMRECDVKN VKKALSMVDEMLQWGQELSLPVLSALVRLVCRS TCHVKAIPDFLDKVFQLFRKLDHETLNFVVEALGK NRFVDQCRLFLHEMYRRHLPIFSATYTASLVSLC QFGRKSDLIYCWSLAGKDNWSPEWDDFIIILENL CRRRMLKEALELFDMMLKVNPGLGTQICDVFLEK LSLNGFTGVGDILVYELKRKNFELDCSAYSHLIKG YCLEKRFSEALIIFDKIRSDSSALSMDALLLLIPQL CKVNRLKEAVALKDAGSKDHPYTLFSLHSALIIGC CRAGKVAEADGLLLGMLLQNMHPKDEIFNALLHA HCHNSSLKTTEKLICIMMKRGICLSVSSYRDLIYLL CKKGMFYSAQGLKSLVVVRNNIPFSVVYNILIFRL FQTGKHYLVDVLVDELEESGQQLDEVGYNFLIHG FSACKDMKRSMKFFNDMINKGLRPSNRSLRAVIH YLCRNGQVQNALELNNQIESRRWILCPSVHYAIM EGLFLDGRIQQAERFVSQIMEKGSTPETIYYNNLI RNFCRYGRLETAVNLLDQILKNKSIPDASSYDCII

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
300	sp_121370_awkj	#N/A	0.9480288	#N/A	lectin protein kinase family protein; FUNCTIONS IN: in 6 functions; INVOLVED IN: protein amino acid phosphorylation, recognition of pollen; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480), Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), S-locus glycoprotein (InterPro:IPR000858), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: S-locus lectin protein kinase family protein (TAIR:AT2G19130.1); Has 120498 Blast hits to 118775 proteins in 4615 species: Archae - 117; Bacteria - 12995; Metazoa - 44949; Fungi - 10067; Plants - 34445; Viruses - 432; Other Eukaryotes - 17493 (source: NCBI BLink).	MKQFQLCSKPTFIFFFFFFTLSASGADTPGGTESA SNPTQTWSSPNNSFSLHFLSLNSSSFSLSITYDS LPIWSISAVADAAATFSFLADGNLRLLSGAGNTL WQSNTSNSGVSHATLDDSGNFCLKNRSNFVVW STFDHPTDTIVPSQNFTVGMVLSSGLYSFSLNRT GNLTLQWNKDIVYLSLGLNATVSTNLSNPRLGLQ SIGILSLYDDNLSGSVIMAYGSDYAESGNILRFLKL DSDGNLRIYSSSKGSGTSIVRWTAVNDQCQVFG YCGHMGICSYNDTSPVCGCPSENFDPVDPNDSS KGCKLKREIRDCPGSVAMLQMDHTKFLTYPPENL NDVYFVGISACRGNCLSGGSCVGSTILADGSGQ CYIKTSSFVSGYQSAALPSTSYVKVCSPVEPNPP PSSEEGGGGGKGKLRGWVVAVVVLTTLVVLIFL EGTLWWWCCRKSPKFGGLSAQYALLEYASGAP VQFSYRDLQQATLNFKEKLGAGGFGVVYKGVLA NKTIVAVKQLEGIEQGEKQFRMEVATISSTHHLNL VRLIGFCSEGKHRLLVYEFMKNGSLDSFLFATE NDKNGRVLSWENRFSIALGTARGITYLHEECRDC IVHCDIKPENILLDDNYSAKVSDFGLAKLVNPRDH RRTLTSVRGTRGYLAPEWLANLPITSKSDVYSFG MVLLEIISGRRNFEVSDQTDRKKFSNWAYEEFDK GIVKTILDRKIADSDVNMDQVTRAVKVSFWCIQE QPSQRPMMGRVVQMLEGIMEIEKPPPPKAVHEG SISGTSLNMSSSISALSTFATSAHGGPSSSSFQ TPVIEELKELWGMAFPITAMNCLVYIRAVVSVLFL GRLGSLELAGGALSIGFTNITGYSVLVGLASGLEP VCSQAYGSQNWDLLSLSLHRMIFILFLAIIPISFLW INLGPIMIFLGQDQEVTAMAAMYCIFSLPDLLTNT FLQPLRVYLRSQQVTKPMMWCTLLAVIFHVPLNY WLVMVLGWGVKGVAVSSVVTNLNMAVMLGGYV CFKEKKQNQNQMMKMKWWNWGLKGVGVEG LKPLLRLAIPSCLGICLEWWWYEIVTILSGYLPNP
301	sp_126610_yyfw	#N/A	0.9228603	#N/A	2A6; FUNCTIONS IN: oxidoreductase activity; LOCATED IN: endomembrane system; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT1G03400.1); Has 8330 Blast hits to 8293 proteins in 997 species: Archae - 0; Bacteria - 1128; Metazoa - 114; Fungi - 883; Plants - 4883; Viruses - 0; Other Eukaryotes - 1322 (source: NCBI BLink).	MDEYSNGSQPSYDRVKELKAFDDTKTGVKGLVD SGVKTLPKIFIRPSEELLKEQKTPCVNLQVPVIDF DGIGKKDRHRYSDIVKEVLDASEKWGFFQVVNH GISSELLDNMIRSVRMFHEQDAEVRKKLYTRDRA KQVLYQSNFDLYTSRYANWRDTLTVDTTCSGHL DPNELPEICRDVMVEYFNQVLKLSDVLLELLSMA LGLGPDRLKEMGCNKGWTAVNLYYPACPDPKLT LGSSKHSDSTFFTLLQDQIGGLQVLHENQWVNV RPIPGALVVNIGDMLQIMSNDKLKSVYHRVTANE VGPRISSAFFLKGILSSPKLYGPIKDLLSEENPAKY REFTISEFQGNFYSRPLDEPGFQYFK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
302	sp_133440_aeew	#N/A	0.946577	#N/A	ankyrin repeat family protein; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT4G03500.1); Has 38664 Blast hits to 17395 proteins in 690 species: Archae - 39; Bacteria - 2692; Metazoa - 19825; Fungi - 3120; Plants - 3302; Viruses - 203; Other Eukaryotes - 9483 (source: NCBI BLink).	MANNIISQLLLAARDGDVKFLTEALATQPADYFL VRSKNKDGQEHYNFIHIAVLNEQSEFLKEALGRL PVSTVHLLCNQASSDFNYNVFHCAALRGNLAIVN LLVKIYKSSSSSSSSSSSSSSSSSMVDLHLHSKPWL TVDVNGKTPFHVALDRGRDDCALEIMSLDKEILC NMVDNKGISPLFLAAQQGCAQVGVKILTSGHSYN TSGEDGSTPLDVVPNCSEEVCKLLLDKHPEMVK HVDENGRTLLHKWVMMGKLWPIQLLLNREKPN MWRKEFINLLCTTEKLTNNNPLHIAACNHNEAVV QVVQLLVEAYIDAYEQGEVLQPSPWKWENKDG DTPLMVSLINKHEELALYFLTVDMENLVIHATKSIL YYAVLNGCNDAAENILITVHPTSFDLMQLKGNGG RNVLHIASNCTERTGDILVEKLPWLINELDEVGKS PLDIAAEVGTAWLIELLLKKDPSLSANAPFAWIEA CKRGHKDAIHAFIDNSTDFRTLCIRRKDSPLHHIQ LKSYWKYKEFLDIPLIQEMKNMPDFTGATPLHRA LERRDIDLVEALLSIDGVFRNIKTKNRKSATEILKE LCKEDYKWDSMCRRIQINPWLRTEFISTGMANC MRYSGSRSQGTTPSAGEMRGTLSVVAALLATLT FAAGFTVPGGLNEDDGEAILAKKAAFLVFILADTY AMCCSIMLVLFGLIWSMVGDRDKSLILIDRSMVILV QSLYGTLVAFMTGIYATISHKSLWAAIIVIIMCSIVA ISSNRAILDKVLDKLIPSVERERRV
303	sp_134300_iads	#N/A	0.934933	#N/A	uncharacterized protein	MIQRFKSAQLLGNVELIPSSNVLGIAAATRDGVAK TMAASAARTRTILLELSTGGRLGSSSAGLN
304	sp_152850_gaix	#N/A	0.941187	#N/A	ATPMEPCRB; FUNCTIONS IN: pectinesterase activity; INVOLVED IN: response to salt stress; LOCATED IN: endomembrane system, cell wall; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pectinesterase, active site (InterPro:IPR018040), Pectin lyase fold/virulence factor (InterPro:IPR011050), Pectinesterase, catalytic (InterPro:IPR000070), Pectinesterase inhibitor (InterPro:IPR006501), Pectin lyase fold (InterPro:IPR012334); BEST Arabidopsis thaliana protein match is: Plant invertase/pectin methylesterase inhibitor superfamily (TAIR:AT1G02810.1); Has 2755 Blast hits to 2700 proteins in 321 species: Archae - 8; Bacteria - 597; Metazoa - 1; Fungi - 199; Plants - 1923; Viruses - 0; Other Eukaryotes - 27 (source: NCBI BLink).	MPASAVCQSTLYPSYCKDVLQSANKTANVYEYG RFSFKRSVSQAARFLKVVNKYLHDSRKLKLTPSA VAALQDCQFLAELNVDFLADSFNAAVNTQENKTL ALTTVEDVQTFLSGILTNQQTCLDGLKSAESAWS FDRGITSPLSTDTKLYSIALSLFTKAWIPKKKKKGS KPPRGKKHPGFFGNNGGRLPGLKMSDNARAVF EASLSRRHERLLGSDNEGDDQVLIYDVVTVSKDG SGNFSTINDAIAAAPNKTEASDGYYLVYVSAGVY EENVSIDKKKRYIMMGDGINQTIITGNRSVADKVF STFQSATFAVTGTGFVGMNMTIRNTAGAINHQAV ALRSGSDLSAFYNCSFEGYQDTLYAHSLRQFYR ECDIYGTVDFIFGNANVVFQNCNIYPRLPLDNQFN AITAQGRTDPNQNTGISIENCVIRASDDLANSNST TKTYLGRPWKEYSRTVYLESFIDNVVDPSGWRE WNGSFALDTLYYGEFNNTGPGSNVTDRVSWAG YHVMNATDASSFTVSTFVFGDEWLPQTGVPYTS GLL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
305	sp_156390_kchz	#N/A	0.993148	#N/A	Adaptin ear-binding coat-associated protein 1 NECAP-1; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: endocytosis; LOCATED IN: membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Adaptin ear-binding coat-associated protein 1 NECAP-1 (InterPro:IPR012486); BEST Arabidopsis thaliana protein match is: non-intrinsic ABC protein 4 (TAIR:AT1G03900.1); Has 463 Blast hits to 463 proteins in 142 species: Archae - 0; Bacteria - 2; Metazoa - 232; Fungi - 69; Plants - 94; Viruses - 0; Other Eukaryotes - 66 (source: NCBI BLink).	MRADEWDVNKWAWEGVLKVISKGEECIIRLEDK STGELYARAFLRKGEPHHVETVIDSSRYLLLRIEE KTGE
306	sp_167450_zthc	#N/A	0.993647	#N/A	uncharacterized protein	MIHNRGRTFSGSVRMQDEPYVTFQVRNYLEPIPL SDFSNASRRGTFRVYASMFEEPAATVQHGFTFR FGRDNATYEVGGSSRQPVHKHIQSSLFCSAPVV GNSQSLPVVSDDFNVVISSANAAMNDPSPNSGIA ATAVISSAMPADRLNALDQRVVEGKPQVFSEGA FSDIPSPQRSFIDSPPLQVSTYQPRMFSESLVLR SVNVIIPPPLMIPKASPVLFSSVVMDKQVVNVSSL KDPPTPTGTKSTGLSPTASLA
307	sp_168240_fyem	#N/A	0.993585	#N/A	MLP-like protein 423 (MLP423); INVOLVED IN: response to biotic stimulus, defense response; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Bet v I allergen (InterPro:IPR000916); BEST Arabidopsis thaliana protein match is: MLP-like protein 28 (TAIR:AT1G70830.3).	MGVFTYTLADITSPVAAPRLFQALAIDNHNLMPKV VPHLVKSVEFVEGDSTTVGCVKQINLPDEAPFKY VKNRVDEIDASNLYLKYTCIEGDVFPETVEVVYE DKYEPSEAGTRCKMVAHYHLKGDNVMKDDDLV KAKEGIQKIFKAVEDHLIANPQVYA
308	sp_171070_eimc	#N/A	0.993111	#N/A	BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT4G21190.1); Has 111 Blast hits to 111 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 109; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MLRLAARYVLQQSSQLGRVSCQSPKIWFSTMHQ TEGQSQVTSNDFGKENSNDHSLFLGQKITRKEK TRFLINALFDLEDSKDAIYGTLDAWVAWEQNFPIA ALKNVLLALEKEQQWHRVIQVIKWMLSKGQGTT VGTYQQLIRALDMDHRAEEAHNIWMKKIGNNLHS VPWQTCHLMISVYHRNNMLEHLVKLFRRLEAFH RKPPEKSIVQKVADAYELLGLVDDKERLLLKYADL LSKNRDESKMKSNRVKSRKKETSSGSRSDVANG EKA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
309	sp_183850_nuza	#N/A	0.931259	#N/A	DOWNY MILDEW RESISTANT 6 (DMR6); FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity; INVOLVED IN: response to fungus, response to bacterium, flavonoid biosynthetic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT4G10490.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEKFVSSWSDGKTLPENYIFPPEQRPGKLILPTS NYVTPVIDLGKLDGPDHSIVVHQILQACQELGAF QVFNHGVARRLMDDARQMLDEFFKLPAEDKSAL ITSDPNSKFKMYTSTVNYDEEKIHLWRDAFQQVC TPLEECIPHWPAKPAKYREIIEEYTRKVNELGLRIL DLICEGLGIEAGYFANEQSQHIELIVNHYPDDEWV GVDPIPEAFVVHIGNIMQIASNGKLLNGDHRAVTN AYEARESAALFFMPTKDCFIEPAKALITAENPPLY RGFRYHEYLAAHNTARRLKDPDLLLKPFRI
310	sp_202780_mzjw	#N/A	0.921703	#N/A	polygalacturonase inhibiting protein 1 (PGIP1); FUNCTIONS IN: polygalacturonase inhibitor activity; INVOLVED IN: signal transduction, defense response; LOCATED IN: cell wall, plant-type cell wall; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: polygalacturonase inhibiting protein 2 (TAIR:AT5G06870.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSNHNKISILSISLLLLSSTISAEKCHPDDKKTLFKI KQAFHNAYLFASWTHDTDCCEWYLVKCDEVTN RIISLSVRDDDEVAGSIPNTVGDLPYLEYLGFTNL PKLTGSIPQSISKLKNLKSLWISHTNLTGSIPDFVG RLKTLTYINLSVNKLTGPIPPSLSSLHKLSALFLEN NQLKGTIPNSFGGFKNNPDFYLKLAKNQLSGPLP KYLGAVHFSVLDLSRNMFTGDASILFGVNKGLQT LDISRNKFSFDLTKVGLPNNLNSLELSHNMIYGSI PPTIAKVSYLQQFN
311	sp_204930_xign	#N/A	0.958515	#N/A	uncharacterized protein	MNMDKGSRKTKLTWKVVECPRSQQVWVVESG YYVMKYMDQTVSLVSKKERDF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
312	sp_004340_zdts	0.927736	#N/A	#N/A	actin binding; FUNCTIONS IN: actin binding; INVOLVED IN: cellular component organization, actin cytoskeleton organization; EXPRESSED IN: shoot apex, stem; CONTAINS InterPro DOMAIN/s: Actin-binding FH2/DRF autoregulatory (InterPro:IPR003104), C2 calcium/lipid-binding domain, CaLB (InterPro:IPR008973), Tensin phosphatase, C2 domain (InterPro:IPR014020), Actin-binding FH2 (InterPro:IPR015425); BEST Arabidopsis thaliana protein match is: Actin-binding FH2 protein (TAIR:AT5G07770.1).	MLKKSKIKSFPCLFVGSAALRPLLPPLLPPPFAPS FLAAPLNPNNFTGDASGQLHRSPLLSSLQRRCA VTNAPTPHLHLAPYLVLGRQSERIHNLAFHATALE LVTDFKKSLNTVNSACDEVRNSGNLKEIMQLILNL GNLLNSGTARGDYVQHKNFTSEFTPYHSSAKGT HHIRHLWGYGL
313	sp_008240_esex	0.350539	#N/A	#N/A	XF1; FUNCTIONS IN: squalene monooxygenase activity; INVOLVED IN: response to water deprivation, sterol biosynthetic process; LOCATED IN: endomembrane system, integral to membrane; EXPRESSED IN: 29 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Squalene epoxidase (InterPro:IPR013698); BEST Arabidopsis thaliana protein match is: squalene epoxidase 2 (TAIR:AT2G22830.1); Has 1994 Blast hits to 1990 proteins in 731 species: Archae - 43; Bacteria - 1249; Metazoa - 112; Fungi - 225; Plants - 178; Viruses - 0; Other Eukaryotes - 187 (source: NCBI BLink).	MVVIDQHIVGIFVVSLLGFVLLFSIFRDVKSKSKSK ANKKKDGRNSVGVQTEVARSSDDVCNVSDGPD VIIVGAGVAGAALACTLAKDGRRVHVIERDLSEPD RIVGELLQPGGYLKLIELGLEDCVGNIDAQRIVGY GLFKDGKNIKLSYPLEQFHSDVSGRSFHNGRFIQ RMREKASTLPNVRLEQGTVTSLLEENGMVKGIN YKNKNGEVMKAYAPLTIVCDGCFSNLRRSLCSP QVDVPSHFVGLLLENCQLPFANHGHVILADPSPV LFYPISSTEVRCLVDVPAGDKLPSIGNGEMAKYL KTVVAPQVPAEIRDSFVAAVDKGNIRTMPNRSMP AIPQPTPGALLMGDAFNMRHPLTGGGMTVALSDI VVLRDLLKPLQDLNDAASLCSYLECFYTLRKPVA STINTLAGALYKVFCASPDEAHKEMRDACFDYLS LGGVFSNGPIALLSGLNPRPMSLVAHFFAVAVYG VGRLMIPFPSVKRVCLGTRLLSGAAGIIFPIMKAE GFRQMFLPATVPAYHRHHHPVK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
314	sp_011450_qrpi	0.853548	#N/A	#N/A	disease resistance family protein / LRR family protein; INVOLVED IN: signal transduction, defense response to fungus, defense response; LOCATED IN: cell wall; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat transmembrane protein kinase (TAIR:AT4G20140.1); Has 135121 Blast hits to 32840 proteins in 1181 species: Archae - 56; Bacteria - 8201; Metazoa - 34800; Fungi - 1569; Plants - 79859; Viruses - 2; Other Eukaryotes - 10634 (source: NCBI BLink).	MAAPANTYPSPPLSSTTSTVVAWATTDSDDSLSC LRTGSSPRTQIIRDEESDFIAHPTRTSLSMLIQMIR AQESEFSASPTTRRPSATGEIRCSCPHFPAENSD NGGNAGGKKPVRNKKSSNGPARRKSLIAFRKAE KDEEKRALEKIQKKVEEDKAERRRKLGLPPKSKG SQQHVISNCLAGGTHESIHCIHDEREALLQFKQGI QFDRCGLLSSWQEQASQHHLECCQWEGVHCN NHTGHVTMLALRGVAGDQPCLEGTVSTSLLKLK HLKYLDLSLNDFGGEPVPSFIGSFTNLQHLNLSK AGFSGEIPHQLGNLFNLASLDLSCDFGNSYGCPY ADSLKWLSHLTLLRDISLSGIDLSQTSDDWYQVA NNLTSLSVLHMERCGLSPSILNSISYMNSSSSFSV IDLSNNNFNDTSIFKWLFNLKGIGSTLVHLDLSDN KIPGPIPHNFGKMLSLSYLDLSSNLLQGPIPSAFW NMRSISYLSLATNQLEGPIPDTISNLKHLSHLILSG NAFQGRGNFPKSLGSLCNLQDIRFNDNNLDDEL STTIQALSGCVRKSLVSLELSQNRIWGSIPDIISTF SLLDGLILNTNRLNGTISPGIGQLQMLEILDLSYNS LKDTISHNHLSNLSRLSILDLSNNPDLVVHISPDW VPPFQLTSLILESCKLGPSFPKWLVTQTNLSYIGI SNAGISDTIPTSFWNSLSPSNLVYLNMSYNMISG VLPNLSVTLLKLPIVDLSSNILSGAIPSFLRNATKL YLNNRFSDLSPFFCPKQKTSLTFLNLANNLFSG EIPDCWMYFDQLLVLHLENNNFSRKLPASIGALTL LYTLHLRNNSLSGELPKSLQNCTSLVILDLGYNTL TGKIPQTIGHSFENLRVLSLQSNNFVGELPLSLCQ LSQLQILDVSINHISGILPKCLYNLTAMTGTEDGLP DLYVDNVYNYYYSEYARLMWKRKEQRFQRSM GQVIAIDISDNELEGHIPEGISSLTGLIFLNLSRNHL SGNIASKVGQLVSLEFLDLSNNHLSGEIPTTLSEV ASLEILDLSNNNLSGKIPVGTQLQSFDSSSYMGN PGLCGLPLPKCSGDQPPANNVQEDDAFSEQNED

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
315	sp_011710_wkjp	0.933118	#N/A	#N/A	BRU6; CONTAINS InterPro DOMAIN/s: GH3 auxin-responsive promoter (InterPro:IPR004993); BEST Arabidopsis thaliana protein match is: Auxin-responsive GH3 family protein (TAIR:AT1G59500.1); Has 1707 Blast hits to 1543 proteins in 239 species: Archae - 1; Bacteria - 597; Metazoa - 55; Fungi - 2; Plants - 674; Viruses - 0; Other Eukaryotes - 378 (source: NCBI BLink).	MGIHQENGDSNNALKFIEDVTSRCDEVQCQVLA EILSQNAEVEYLKRYGLPKGCVDREKFKSKVPVV MYEELKSDIQRIANGDFSPILCARPFTQLFSSSGT SGGEPNLIPATDQEIPRRHQQYKVLLPLLSQCFP GMESGKLLCGFYQRHEVVRVDSHFASTLVVVIHF LRDHYSELCDDMSSGTLSSKISDMALRAHMAEH VVKQPEPELAKLVYNACKGENWEGILKKIWPNIK YLDIITGSMAQYIPILNFYSGGVPIVSLKYTCSEGA FGINLNPLCDPYEVSYTLIPNMAYFEFLPLNTSSS DREQSSDQLVDLADVKVGHEYELVVTTYDGLYR LRVEDVLSPTGFYNSTPQFKFVRRNNVLLSVHAE KTTESELQNTMEKVYDVLQNFGTIRVVDYTSYAD VKTIPGHYVIYCELSTTTSDATNKLGLDGVMEKC CLTIEDSMSYVYRYYRVHNPAIGPLEIRVVKKGTF DKMRDFAISRGASLGQYKVPRCVKFIPLLELLNS GVISTHFSPSLPHCSL
316	sp_011890_wuzw	0.931111	#N/A	#N/A	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity, iron ion binding; INVOLVED IN: oxidation reduction, flavonoid biosynthetic process; LOCATED IN: peroxisome; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Isopenicillin N synthase (InterPro:IPR002283), Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT4G16765.1); Has 6583 Blast hits to 6566 proteins in 839 species: Archae - 0; Bacteria - 934; Metazoa - 52; Fungi - 818; Plants - 3817; Viruses - 0; Other Eukaryotes - 962 (source: NCBI BLink).	MGEEDEDSLIVNLPIIDLLAPDRVSTAATLRQACIE VGFFYIVNHGVDKDLLGRLFEESKKFFDLPVEEK MKLDRKEIRGYTPLYSEKLDTSLNTKGDSKESFYI GPLHDELCNLNQWPSEELLPNWRATMEAYHKQI LNAGKSLLSLLAPALNLNENFFENVGALDQPMPV LRLLHYPEGDNNCNNDEVLGASAHSDYGMITLLV TDGVPGLQVCREKDRQPRIWEDVPHVEGAIIVNI GDITERWTNCLFRSTLHRVVPKGRERFSAALFLD PPHNFSVECLETCCSESNPPRFPPIRYNDYLKER YRVTYGA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
317	sp_012520_wxgw	0.3440679	#N/A	#N/A	GDSL-like Lipase/Acylhydrolase superfamily protein; FUNCTIONS IN: hydrolase activity, acting on ester bonds, carboxylesterase activity; INVOLVED IN: lipid metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Lipase, GDSL (InterPro:IPR001087); BEST Arabidopsis thaliana protein match is: GDSL-like Lipase/Acylhydrolase superfamily protein (TAIR:AT5G45670.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MMRGAVAVLTLLIIITIVRCEPQVPCYFIFGDSLSD VGNNNDLVTKAKANYPPYGIDFPGAIPTGRFSNG RVFVDRITELLEFDNYILPFSSAKGEEILKGVNYAS GAAGIRYETGQHLGERISMYDQIMNHNETVAKIK EIGKTSAVPKLEECLYTVNIGSNDYINNYFKPGYY YISSSLYNPDEFAEELINQTYAANLKILHLSLGARKIA VFGLGLIGCSLGEIYTFGTNGSLCVDKINAAVQPF NQKLVNMVDDLTMELSDLQLTYIGSAASTDTGSA IADAPCCKVRDDFQCIEFEAACEGRNNYIFWDGF HTTEAVNVLTGEIAYSALDTKHIHPVDIKTLVSIDR ESMINQVKNSMSSLVSSA
318	sp_013260_kgki	0.3647837	#N/A	#N/A	Lung seven transmembrane receptor family protein; LOCATED IN: endomembrane system, integral to membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Transmembrane receptor, eukaryota (InterPro:IPR009637); BEST Arabidopsis thaliana protein match is: Lung seven transmembrane receptor family protein (TAIR:AT5G18520.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MALHHSATLTFLIFVALFSLSTAEIRTTKVKSDNR PIIPIDEFGFTHTGRLELNVSGITLSTRNKLDLSKL GFFLCTRDSWLHVLEQLEESEISCALDSHLVKKV YTFNSLPTTDPDSFGSLYVETDSDQYTLVFANCY PENVQISMNIRSAMYNLDGKTRTRDYLSAGRTIL PRIFYFFSFIYFGIAVFWTYVLYKKRLTVFGIHFFM LGVVIMKALNLLCEAEDKSYIKRTGSAHGWDVLF YIFSFLKGITLFTLIVLIGTGWSFLKPYLQDREKKV LMIVIPLQVVANIAQVVIDETSPFNQDWVTWKQV FLLVDVICCCAVLFPIVWSIKNLREAARTDGKAAV NLMKLTLFRHYYIVVICYIYFTRVVVYALETITSYR YLWTSVVAAELATLAFYVFTGYKFKPEAHNPYFA IDDDEEEAAVEQLKLEDEFEL
319	sp_014380_pjhc	0.3553113	#N/A	#N/A	uncharacterized protein	MEDDQPPTFEELLLLEELLESGQLSDDDLLDVD IPTMDFLSMEQQMNQFRDDEASQAEVWYLSDD DNNVDYEAMHNVYDEHIHSDIISSYFQNMQPSET SNTNRNHGKKTFELSDNQRAVMVAKLMLLANKD MTLPWGVAKKLANEYDTVEVPDDLLTSIRQYMQ EGQININEFRQRVNEMGVRVEDPYQNGRHQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
320	sp_017120_efpt	0.9376831	#N/A	#N/A	F-box family protein; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364), F-box associated domain, type 3 (InterPro:IPR013187), F-box associated interaction domain (InterPro:IPR017451); BEST Arabidopsis thaliana protein match is: F-box and associated interaction domains-containing protein (TAIR:AT5G52610.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSDIPPEVINKILERLPVKSLLRFKCVSKTWYSLIS NRNLEFAINNLKHCLSSNDNTLLAIFNDNWSLQS NEFDIKDPFSPVTRFDLGIPSEYDSEDYCFNARIV GSCNGLVAVVLDGCLCLWSLKTELFRDTNWVV MKEYGVKESWAPLFSLSEIDIFKPLPMPLIAFSKD EFLVRQDVGYVVLRTLPILYADSNSLQSLMYIAYL VLRTLHILYADNLTVPSPGVDALGCHCCVGSLVL VPNGVPQNRQ
321	sp_017920_yrqg	0.9323119	#N/A	#N/A	uncharacterized protein	MGLKKHVLSRPPPTPSPSPSDSMTRNKNGYVDR VSVQRPPTYMDVSANSSPMERSTPSFHVERSTP SIQAESEIMPNYNSDEPDKSVFKGSVQPPDIWTD EKGFRYYVKLNEFCQPLRKGGSILVRFLGDIAKQ DSLCPVGIKTRAVNKKLKSNIVTMVREHFVIPDG LVVDKAIVRAGKSWKNHFYILKKQYFDPVNKTL DQNYDSIPDGISSSSWNDLVNYWFSLKGTKEKK GEYRELTFYKSVYAKEDGRFKEGTISRQFMEDA NNKVQENLASSSTSKSKVEIENE
322	sp_020030_fqha	0.9323119	#N/A	#N/A	uncharacterized protein	MARETVTVAHHNFHFLLTEKSKVVVSTRIQFFYK TTVTIAPLFKASLGP
323	sp_020300_mcup	0.9263335	#N/A	#N/A	Phototropic-responsive NPH3 family protein; FUNCTIONS IN: signal transducer activity; INVOLVED IN: response to light stimulus; LOCATED IN: plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: NPH3 (InterPro:IPR004249), BTB/POZ fold (InterPro:IPR011333), BTB/POZ-like (InterPro:IPR000210); BEST Arabidopsis thaliana protein match is: Phototropic-responsive NPH3 family protein (TAIR:AT3G50840.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAAEKSGTSTPSKGQAWFCTTGLPSDIIVEVDN MTFHLHKFPLMSKSKKLHDLIREQEETKNLPTNK RRQDTQNDDVESKSEEDDEEEDENLLEQGQCH VSLTDFPGGSEIFETAAKFCYGVKIDLTAAKVAPL RCAGEVLEMTEEYSEDNLISKTERYFANTVLKNIK DSVKTLKSCEKLPPLSENLGIVDRCIDAISSKASLY DPTLFGWPVNDVVEESRKKKKKKNAAAPPSKVSD AWVEDLAQLSSHLFNRVISTMRMKEVNPEVIESC LIHYAKRCFPGISRSNRKPTATAPSMMASEAEQR QLLETIISNLPPENHTRSPQAAATSFLFGLLRTANI LNASETCKSALEKKIGAQLNQATLDDLLIPSYSYL TETLYDVDCIERILSHFLANHEEREIEGEGDHQTL GNVRTPAIMHVGKLIDGYLSEIASDANLKPGKFIQ LALSPPEKARLYDDGVYRAVDVYLKAHPWLSEE ERDKVSGVMDCQKLTLEACTHAAQNERLPLRAV VQVLFFEQLQLRHAIAGTLITADIAPSESTPPSLLG HDEEEERDRGGGGGSGGGGGGGTWKEAIKENQ VLRLDMDSMRTRVYELERECSSMKKVIEKFDKD KVKVGVVNSPKRWRSTMS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
324	sp_022650_pgkf	0.9285531	#N/A	#N/A	amino acid permease 4 (AAP4); CONTAINS InterPro DOMAIN/s: Amino acid transporter, transmembrane (InterPro:IPR013057); BEST Arabidopsis thaliana protein match is: amino acid permease 2 (TAIR:AT5G09220.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MNFDDAGRIRTGTVWTATAHAVTAVIGSGVLAVP WSVAQMGWIFGPLSLFGFAVITYITARLLCDCYR TPDPVKGQRNHSYMDAVRATLGKKEVLICAFAQ YSLLWGTMVGYTITAATSMLSVKRTNCYHTKGA EADCQPSSNSFLLIFGAIHVLSQLPSLEKVTFLSV IAATMSFCYSFIALYLCVAKFVSNPVVRGTLTGVE IGINGGGLSLSSKLWQSSQALGNIAFSYTYAQLLI EIEDTLKSPPAENKSMKRATLYSIALTTTFYVSLG CMGYIAFGNEAPGNVLTAFHEPFWLVDLANIGVV IHLTAAFQVFAQPIFASYEKWLAIRWPESPLIHNT YTIQVPRFKALSYQFTLCRLLLRPLVVVVTVLTAM MFPFFNAVLGLLGSIAFFPLTVYLPVRMYMTQVK LNKGALKWCMLQGLSLLALFVSLISAVGSVADIA DRLKNTKLFNANL
325	sp_026020_gcng	0.9402884	#N/A	#N/A	basic helix-loop-helix (bHLH) DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Helix-loop-helix DNA-binding domain (InterPro:IPR001092), Helix-loop-helix DNA-binding (InterPro:IPR011598); BEST Arabidopsis thaliana protein match is: basic helix-loop-helix (bHLH) DNA-binding superfamily protein (TAIR:AT3G56770.1); Has 2073 Blast hits to 2066 proteins in 145 species: Archae - 0; Bacteria - 2; Metazoa - 255; Fungi - 6; Plants - 1802; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MPHHFNATNTTSSSMHLQELPANCYALEFQEPF EGFAEFEDKASAADSSHNHRQAEKRRRDRINGQ LARLRRLIPKSEKMDKAALLERVVEQVKDLKRKA SEISKIFTIPSEIDDITIECHRKENHPNNNNNNNNN NNSNSNIMFMKATLCCEDRPDLIPELIRELKDLNL TVIGADIGCLGGRMKSILLLYYKLDSNNDGVVCNV SAIKQSLMGVLTRIISWSSASTSSFRVTSKKRQRFF FSS
326	sp_026820_nzto	0.9223119	#N/A	#N/A	Ankyrin repeat family protein; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT1G34050.1); Has 20807 Blast hits to 11728 proteins in 551 species: Archae - 16; Bacteria - 1568; Metazoa - 11429; Fungi - 1235; Plants - 2476; Viruses - 139; Other Eukaryotes - 3944 (source: NCBI BLink).	TEETFKKEDIESMVKAKDEASHWLDKDGLTPLFR AFQTGNLSTATELCTISSEAAGTCDANGQTFWH LLVHHPSGYFVFRLLQNEVLRDLLKLKDKKGNTP LHLAIEGNRYDLVKDFLDIWERERELLPKGKDQW LVNLLQIQNKAGKTPADLIRESPSLPPNVSHP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
327	sp_027800_iwai	0.923119	#N/A	#N/A	indole-3-acetic acid inducible 32 (IAA32); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: response to auxin stimulus, embryo development ending in seed dormancy, pollen development; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Aux/IAA-ARF-dimerisation (InterPro:IPR011525), AUX/IAA protein (InterPro:IPR003311); BEST Arabidopsis thaliana protein match is: indole-3-acetic acid inducible 34 (TAIR:AT1G15050.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MESNMASFLLNTSALDLGLSLNTIHTQACDPVDQ RRQGMQLHSTIIPGGYNNNHNEMEWPYLKSPKC SKSNPDDCGEDLEGVESKERWPYVKVNMDGVIV GRKICLVDHLGYPSLALQLEDMFGGFTPYGVRLF QRGSEFMLFYKDREENWRSAGDVPWKDFVDCA KRLRVVRKNRYLLS
328	sp_032290_xsaw	0.876432	#N/A	#N/A	uncharacterized protein	MGTDSVVAHYPSVFSSDCSYQGDIGSELDEGVA VPPVGMSYRRDMDVFDFSVPAEEVMEVLSDYVF R
329	sp_037830_teft	0.923119	#N/A	#N/A	uncharacterized protein	MDEA
330	sp_039740_fjzg	0.923119	#N/A	#N/A	uncharacterized protein	MVSNAMAILRLETRRRRRGGGPYRSGEGRSGRI RVGDLIQN
331	sp_040000_ztxd	0.926624	#N/A	#N/A	Eukaryotic aspartyl protease family protein; FUNCTIONS IN: aspartic-type endopeptidase activity; INVOLVED IN: response to salt stress; LOCATED IN: cell wall, plasma membrane, membrane, plant-type cell wall; EXPRESSED IN: 6 plant structures; CONTAINS InterPro DOMAIN/s: Peptidase aspartic (InterPro:IPR021109), Peptidase aspartic, catalytic (InterPro:IPR009007), Peptidase A1 (InterPro:IPR001461); BEST Arabidopsis thaliana protein match is: Eukaryotic aspartyl protease family protein (TAIR:AT1G03230.1); Has 1398 Blast hits to 1391 proteins in 46 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1396; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MASGPFLLSFLFIFSAFSFHFSNANLVLPVVKDNT KSQYYTTFLVGTRPSTRVYLVIDLANQWSWFAC DLYGYNTNSHEVPCNSPKCERYAPGTFCSGNAC GALAYTPWSNGLFAGDLYEDTVSLYKQTQQNSL AQLSKVYLPKLPFTCADKGPLKGLSPRTKGVVTL ARVPVALHSQISSTFKVARKFALCLPSSTSLGYN GAIYFGGSPDYLSKSLITTPLVINPNSTAPIYSEGD ASVEYFIKVRSIKIDGTPIRFRSSLLSFDIKEGVGG TKVSTMDPYTVLHSEIYKTLVNVFVTKATAMKMT RVASVSPFGTCFSSKSVVDNAKTGPAVPIIDIIMD GKSPLLHNKNARWRIYGANSMVKVSDTVMCLGF VDGGSKPTTSIVIGGKQMEDNLVEIDLETSKLGVT SSLLRMGTNCSKFKGI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
332	sp_040030_zdic	0.922837	#N/A	#N/A	Eukaryotic aspartyl protease family protein; FUNCTIONS IN: aspartic-type endopeptidase activity; INVOLVED IN: response to salt stress; LOCATED IN: cell wall, plasma membrane, membrane, plant-type cell wall; EXPRESSED IN: 6 plant structures; CONTAINS InterPro DOMAIN/s: Peptidase aspartic (InterPro:IPR021109), Peptidase aspartic, catalytic (InterPro:IPR009007), Peptidase A1 (InterPro:IPR001461); BEST Arabidopsis thaliana protein match is: Eukaryotic aspartyl protease family protein (TAIR:AT1G03230.1); Has 1398 Blast hits to 1391 proteins in 46 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1396; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MASSNRTFSLTFLVMLSALSFHFSNAHFVLPVVK DNTNSKYSYYTSLLVGNHHPSRVNLVIDLSNKWS WFTCDFFGYNTTSHGVPCHTPKCDRYAPLAPCS DGICVVDAYTPSNFIYSNNLLEDTTKVYKQIQQN SLTQLSNVNLPNFPFTCGDNVPLKGLSPHTKGVV TLARVHVSLHSQISSTFKVARKFALCLPSSSPTSS GYNGAIYFGGSPNYLSKSLITTPLVINPNSTSPIYS PGDSSVEYFIKVKSIEIDETPIRFRSSLLSFDIKEAV GGTKISTMDPYTVLHSEIYKTLINVFITKATVMKMI RVASVAPFGACFSSKSVVNDAKTGPKVPIIDIMD GKSPLFARWRIYGANSMVKVSDTVMCLGFVDGG LKPTTSIVIGGKQMEDNLVEIDLETSKLGVTSSLL HLGTSCSKFKGI
333	sp_041370_tnfi	0.922536	#N/A	#N/A	Expressed protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MAVVTQEPILCRLDRLDNLMRHLENRGSSNNNN NNNNSNRSSSPSTPSSGTARNSDGHASSLDFFS PRSRDKLHCRPIDDVIVETEHKGTLLDRVSLIEHR LLKLCMQLEQEMDAEKQDKHSQDENKSRSPQT HKKSGLKQFVKSCVKGNKHKPNK
334	sp_042210_agke	0.934436	#N/A	#N/A	GATA transcription factor 1 (GATA1); CONTAINS InterPro DOMAIN/s: Zinc finger, NHR/GATA-type (InterPro:IPR013088), Zinc finger, GATA-type (InterPro:IPR000679); BEST Arabidopsis thaliana protein match is: GATA transcription factor 9 (TAIR:AT4G32890.1); Has 1635 Blast hits to 1587 proteins in 195 species: Archae - 0; Bacteria - 2; Metazoa - 169; Fungi - 549; Plants - 842; Viruses - 0; Other Eukaryotes - 73 (source: NCBI BLink).	MESFDKAARFMDELLDFSSDIGEEESDDEIPGNS SKSHHFLHHHRLNHRRHNRQSPPFAEFVEEELE WISNKDAFPMVESLVDILPNHPRIASSYNTNHQS PISVLDNSNSSSSSSANSHLTHGSLNNNNINSNG GRGGNMMLINLSSDFKVPGKARSKHGRKFRKD MLGSQFKCFVYEKKGARKLASSVSASSKSSSTL GRKCLHCGSEKTPQWRAGPSGPKTLCNACGVR YKSGRLCDEYRPASSPTFSSDLHSNSHRKIVEMR RQKMPLTGSGSNSLEIG
335	sp_047220_sodq	0.933119	#N/A	#N/A	uncharacterized protein	MTPGGSPSKSIGFGRQGHGDGGGGGGGAGA GGGRGRRSDGWVWPPRPRIKNSVLHFNNQGTC SCASQIPRQQQQQQQQQQQTISLTMPSSISYSVY HNVSAPPVSIQQGNNPRPLRFLFCKELTYSDVKK RNKILPRVIVASKERNEVLNGECSVPNNGDLGM GLMTSSYSPSACNYVYLNQLSDFCVVDEMEEVA TPRTPYPYGNATWTATIVSFDLPPLQTFNLFDFPI DNMSWT
336	sp_048110_xhda	0.944873	#N/A	#N/A	uncharacterized protein	MFKMGKPKIAVCFIILFSLSSISISDGACSPSEDN NKYPVIGILSHPRRNNG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
337	sp_048300_qhsr	0.923119	#N/A	#N/A	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; FUNCTIONS IN: lipid binding; INVOLVED IN: lipid transport; LOCATED IN: chloroplast thylakoid membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Bifunctional inhibitor/plant lipid transfer protein/seed storage (InterPro:IPR016140), Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor (InterPro:IPR003612), Plant lipid transfer protein/hydrophobic protein, helical domain (InterPro:IPR013770); BEST Arabidopsis thaliana protein match is: Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (TAIR:AT5G46890.1); Has 782 Blast hits to 776 proteins in 62 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 782; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MASKAFASTSLLLCLNLLFFTLVSSTYAPNCPPSS GHHGGHHGGHHGGHHGGSPSTPTTPIMPPAQA KCPIDALKLGVCANVLGDLINVVVGTPPKTPCCPL IQGLVDLEAAICLCTAIKANILGINLDIHLSLTLLLNY CGKGVPSGFQCPY
338	sp_048730_mfei	0.923119	#N/A	#N/A	uncharacterized protein	MEGKQKCRTFCPKGAWRLPPILPYWDKIRKLDD RVKELDSLKVSSNSAPSTEATKKCAIMSHPSPEN FPGKTNVVESLEKVVSVQPCESSACSNTAQDIYE VPICSYIDSSNIKKESYNEPLLSPESAKSSKKSDS KDKIDWEKKPLLSPGSSNESPKLKDRKTDKPLLF KHPTLGGVADFELQQFNRRLKQLEDQRNGKLDL AWGEIWKSDVKELLISPESAKSSKTNDSKDKIDW EKKPLLSPGSSNESPKLKDRKTDKPLLFEHPTLG GVADFELQQFNRRLKQLEDQRFVTRGHETLKKL KKKN
339	sp_048870_qkcd	0.923119	#N/A	#N/A	Tetratricopeptide repeat (TPR)-like superfamily protein; FUNCTIONS IN: peptidyl-prolyl cis-trans isomerase activity, binding; INVOLVED IN: protein folding; CONTAINS InterPro DOMAIN/s: Tetratricopeptide TPR-1 (InterPro:IPR001440), Tetratricopeptide-like helical (InterPro:IPR011990), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734), Peptidyl-prolyl cis-trans isomerase, FKBP-type (InterPro:IPR001179); BEST Arabidopsis thaliana protein match is: rotamase FKBP 1 (TAIR:AT3G25230.1); Has 3730 Blast hits to 3559 proteins in 279 species: Archae - 0; Bacteria - 41; Metazoa - 2106; Fungi - 376; Plants - 567; Viruses - 0; Other Eukaryotes - 640 (source: NCBI BLink).	MKEAGNLLFKQGDIEDALEKYGYAGIILGCFQFEE NDDRAEFFDLANCILLNSAACFSRKREFEQVGKI CSVILEFNPNNVKASYRRAMAAIELGRSDLAYPD LLMASHNDPRNTEVRKKLEEVKQSREKEVSGKH SLGDVPIGLGLGLSPPRKKSKGELVKEHTIDQVQ LHEKEHIMSEKIGCDVRLGLNSTVSEPVLSIMKDK HLSESSSVVEENLMSAVLEGAEEMMEDVLSVSN SENSKPVEPNYRVVNRLRSGSSLSISKKDYQLLL KGKTIQYFNSKVASPMTIRVIGGGLDNTMKRNYE CHEKRLEDSFRVDMPQHSSIVSGLNEELENQRE ERLRLDGHYNLFDSNKESESESMSVTTENYHHIG DASEYRDDACFKSTASIGQDEKFDPGVKVRKKD GRCIGRKATRKQVDKKENAIVTHSYPVFSSIKVEK KRKVVTPPSDCIQRPRKKVLYASLTNSIESYASSF ESCASSVESCISSDKSCNNKVICVSHEMANSQE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
340	sp_048940_weag	0.922613	#N/A	#N/A	Heavy metal transport/detoxification superfamily protein ; FUNCTIONS IN: metal ion binding; INVOLVED IN: metal ion transport; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Heavy metal transport/detoxification protein (InterPro:IPR006121); BEST Arabidopsis thaliana protein match is: Heavy metal transport/detoxification superfamily protein (TAIR:AT4G10465.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MPKSSRPLSLQTVELKVRMCCTGCERVVKDAILK LRGVDSIEVELSMEKVTVVGYVDRNKVLKAVRRA GKRAEFWPYPNPPLYFTSANNYFKDMAHDYKET YNYYRHGYNVGDKHGSLPISSRGDDKVSNMFND DNVHACTLM
341	sp_050470_ptpo	0.923319	#N/A	#N/A	ent-kaurenoic acid hydroxylase 2 (KAO2); CONTAINS InterPro DOMAIN/s: Cytochrome P450, B-class (InterPro:IPR002397), Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 88, subfamily A, polypeptide 3 (TAIR:AT1G05160.1).	MFGKPCVILITPESCKKVLSDDATFGPGPPESITK LMGKKAFHGITQEEHKRLRRLTTAALSGQEALSI YLKHIEEITINSLKEWAGMSRPIEFLTEMRKIAFKV IMFVMVGDTSTDHFIMEALEKEYTTLNHGLKSMAI DLPGFAYHNALKARKNLQKIFQGVVDKRRIEETM ANKQEEKKDMLDMLMGIEDENKSKLSDEEIVDLII MFLNAGHESSGHIAAMWAVVLLQDHPEVFQKAK VINTVFGNQVDVEM
342	sp_053050_hrne	0.92249	#N/A	#N/A	nuclear factor Y, subunit C3 (NF-YC3); FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus, intracellular; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Transcription factor CBF/NF-Y/archaeal histone (InterPro:IPR003958), Histone-fold (InterPro:IPR009072); BEST Arabidopsis thaliana protein match is: nuclear factor Y, subunit C9 (TAIR:AT1G08970.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MDSHWSLNHSDNKFSFVDTTTRYINQQQQQQQ CQQQKQKPHDSASGGAGGSSPLLQQQLRGFW QSQIQESKQDVDFRNHNLPLARIKKIMKSDDNVR MISADAPVLLSRACQMFIQDLSLRAWNSTQEGK RKTLQKTDISAALSTTDLFDFLVDILPREKPLISAT MLPPVNSASSSLGVICDHQLSQHQGYSVGEGTG LTMGLQMHQLVQDQDDDQLDGGCNNSSNYIHRG WPNYLSHHSAWPPP
343	sp_053090_gkfc	0.962537	#N/A	#N/A	uncharacterized protein	MDGAKNAAETVKTNVTEQTDKIDTDKISKTAGEM AETTKTHVGDALEKAGDVVKPEPPKEESLMDKA KGLFK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
344	sp_054370_dsqe	0.953539	#N/A	#N/A	embryo defective 1075 (emb1075); FUNCTIONS IN: pyridoxal phosphate binding, carboxy-lyase activity, catalytic activity; INVOLVED IN: cellular amino acid metabolic process, embryo development ending in seed dormancy; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Pyridoxal phosphate-dependent transferase, major domain (InterPro:IPR015424), Pyridoxal phosphate-dependent decarboxylase (InterPro:IPR002129), Pyridoxal-phosphate binding site (InterPro:IPR021115), Pyridoxal phosphate-dependent transferase, major region, subdomain 1 (InterPro:IPR015421); BEST Arabidopsis thaliana protein match is: glutamate decarboxylase 2 (TAIR:AT1G65960.1); Has 3259 Blast hits to 3251 proteins in 995 species: Archae - 187; Bacteria - 1878; Metazoa - 518; Fungi - 211; Plants - 249; Viruses - 11; Other Eukaryotes - 205 (source: NCBI BLink).	MAGCHALPQEIHQDFLAVAHITSNGDTSGEARKH LIDNNGEREIVVRQRRRVHTTDLEINEQFGANGE AKKSCDVNGKREIVVTRTDLEIKEQPGAANGNLI GEKEAYMEDVLVKHIETWMDKTKYHLGYPGNLD FNFDDLGDRLQQFLLNNCGDPFIEGNYGMHTKD FEVAVLDWFASLWEIEKDQYWGYVTNGGTEGN FHGLLIGREMFPDGILYTSKESHYSVFKAAHMYR MDCIKVDTLVSGEIDYGDLKTRLLENKDKPAIITLT IGTTVKGGVDNIDLVIKTLKECGIPRNRFYIHCDGA LFGLMLPFINRAPQVTFKKPIDSVTISGHKFVGCPI TCGIQIVRKQHINVLAKNVEYIASRDITIMGSRNGQ APIFLWYALNRKGYNGFRKEVQKCMQNARYLRD RLIEVGVSVMLNELSNTVVFERPKNEEFIQHWQL SCQGNIAHVVMPNVTVDKLDMFVDEMVEKRLS QPPCVAFEIGKENGVCSLH
345	sp_055670_rnri	0.923119	#N/A	#N/A	uncharacterized protein	MRSFQPYQHHNHHLYYRPPPSTQPPLLPTPSHY HNPPPFRSHQPPTLPPKPHNYHRQLPTPLPNYH NVLISTSPVDHHSTLSPPAELRVYSPSSPFYSPVS PNYCGAYYYPHCTTTCSPDDDHNAAYSSWSPKY SLPSPVFIPPPSPSPSPPPSPVRVEESSSKSPVE SPPSSPVYSPTCSPSVDYSIAYSPSYYPYTP
346	sp_058510_kipf	0.923119	#N/A	#N/A	uncharacterized protein	MYTFKKQATTASFIGIGVGDNTSTRVGSEVLKPIQ GQKNNSKGILTVDDSDVNIDGAFFDPLFTYPSAT RNNIEGWELSPEPVHEENMDDELSSEDEDMHIQ NGNENFDNHDNEACEDILENVHVDEREISQTIQD VNVEGGGSSFDEKNMNQKKRGRNKCKEIAKLKP DEKLEITFYNNRAIGTNHKVFARHLGIIVRDTNICP VRVRKWDDIGDKEKEHMWSAVTVRI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
347	sp_058560_ccgr	0.923119	#N/A	#N/A	uncharacterized protein	KEGWCWSSVKANNSTNSGGWSTGKANVDGVS NSWRSQKEEKPQPKSNDRSVGDGNFDKDAKEE GLSSWDAKKVERKTQSSWGQPSESKNSAQSSA DHWGFDKSNQPGKSSGWGSEDTNAGKDSEKQ DSWGKSNFGGSGWNKKQPEGGRGWGSSNTG EWKSRKTQNQNQNQNQNQNRPPRGPNDHS PRVALTATRKRMDEEFPTISTPTDRTPTISTPTIATP TDRTPTVSTPTISTPNISTPTDRTPTISTPTDRTPS SAWGKPAQETDNIGWKKNNPQRDSENLEGXFY FNPYCLNPYYFNPYCLNPYYINPTDRTPTISTPTD RTPTISTPTNRTSSSAWGKPAQETDNIGWKKNN PQRDSENLEGTSGWNDKLQKENRSFSKQSQPA SSKDWDSTGNVTTGNTGFGVEKGNEKPWDVAS NVSVKKSTWGQTGGNSWKKNEQDEKDGDPQG LPWGKSHKSSDSWTSGQGNQHPVSQGVXGVS EKQETLFLGTT
348	sp_062440_ojur	0.935833	#N/A	#N/A	glutathione S-transferase TAU 8 (GSTU8); FUNCTIONS IN: glutathione transferase activity; INVOLVED IN: response to cadmium ion, toxin catabolic process; LOCATED IN: cytoplasm; EXPRESSED IN: 7 plant structures; EXPRESSED DURING: LP.06 six leaves visible, LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), Glutathione S-transferase, C-terminal (InterPro:IPR004046), Glutathione S-transferase, C-terminal-like (InterPro:IPR010987), Glutathione S-transferase/chloride channel, C-terminal (InterPro:IPR017933), Glutathione S-transferase, N-terminal (InterPro:IPR004045), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: glutathione S-transferase tau 7 (TAIR:AT2G29420.1); Has 7045 Blast hits to 7029 proteins in 1100 species: Archae - 0; Bacteria - 3564; Metazoa - 831; Fungi - 183; Plants - 2012; Viruses - 0; Other Eukaryotes - 455 (source: NCBI BLink).	MALKLKGIPYDYQEEEPYGNKSPELLKSNPVYKKI PVLIHNGKPIAESQVILEYIDQVWDGYSILPTDPY QRSLARFWAKFIDDKAMPSLWNAMWAKEDEQE KTLKEAKENLEMLEKEIEGKKFFGGENIGFVDIVA NFIGFWLGTIQQVLGKEILTKETYPHLSNWAEEY VNCDIIKQNVPPKERLFTLFKLRHEAGF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
349	sp_063480_wpxr	0.823532	#N/A	#N/A	Protein of unknown function (DUF1442); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1442 (InterPro:IPR009902); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF1442) (TAIR:AT2G45360.1); Has 92 Blast hits to 92 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 92; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAHWSAENATKAYLRTIKIMGIKRGIKEPDVAEFIS ALAGGNNAKLMVEICGSIAGPTTLGLVAAAKQTS GRVVCILPGPKELQASKSQLSRYLNHVEFVIGDP LTLLTNRYNSADFVTVDCNINGHEAIFKTGQKGG ALVVGYNAQHLVPSYNELGGHFLPIGEGLLVSKV PKHHDIRKKSKWIVKVDKCTGEEHVFRITCPQQK VIQA
350	sp_063840_txwp	0.853831	#N/A	#N/A	uncharacterized protein	MRQLMEGETRKRKITNICGEDEEEDEDIKMEKFF ALIRSANDIRDQLLDAEGSSKERAKKAVATAEKK AEVAVAGRAAQRNNEKGFASFATATAAAVDADA VVPNKKAAAAAAAAAAAAAADSAARISRSAAAAA AAEENSKKAADQKNEFKGGRIFLDLNLSL
351	sp_064960_mipu	0.823119	#N/A	#N/A	uncharacterized protein	MAGSSEMKGEESKLSFEEQASLDEIIMNAKKSFC AEDMVNHSVVAENKAEVASGHLIVLGGNLSNLTK GFLTVLDSKLWAGYIIKDYPNPLLVGTSILQLVQV GVSNDGLKAAVLYTGKNIACVDCGWLLAWADSQ GKGKRVYAECGPISKFNCLNWERIEKKLDDPCC PAINPYKASDEQTGTSVIVGLSGVEWPLTYETRD LLAGIISG
352	sp_065850_itki	0.842332	#N/A	#N/A	uridine-ribohydrolase 1 (URH1); CONTAINS InterPro DOMAIN/s: Inosine/uridine-preferring nucleoside hydrolase (InterPro:IPR001910); BEST Arabidopsis thaliana protein match is: uridine-ribohydrolase 2 (TAIR:AT1G05620.1); Has 6361 Blast hits to 6297 proteins in 1364 species: Archae - 75; Bacteria - 4675; Metazoa - 179; Fungi - 233; Plants - 187; Viruses - 0; Other Eukaryotes - 1012 (source: NCBI BLink).	MGTIEESADCLLGESISNGVVYDSSSTYSAKCEKLI IDTDPGIDDCMAILMAFQTQEVDIIGLTTIFGNVQT ENATRNALLLCEIAERPDVPVAEGSPGPLKGGMP RVADFVHGTDGLGNINLLPPKARKIEKSAAQFLV DKVSEHPGEVSILALGPLTNLALAIKKDSSFASKA KRIVVLGGAFFALGNVNPAAEANIFGDPEAADVV FTSGANITVVGINITTQVKLTDSDLHGLRESKGRH TRLLTDMCKFYRDWHVKSDGVYGIFLHDPVSFV AVVRPDLFTYKKGVVRVETQGICAGYTLMDQGL KKWNSENPWTGYSPISVAWTVNVEEVLMYYKS QLTQS
353	sp_070260_rqas	0.823119	#N/A	#N/A	uncharacterized protein	MNREKARKTGEIGKEKPMEAGVGAIEGPIPDIIEG PILEGAIPIPIPISISPFTAIATAATTNNITKLDNLEAI VFVFVTKT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
354	sp_070580_zeiq	0.971136	#N/A	#N/A	RNase H family protein; FUNCTIONS IN: ribonuclease H activity, nucleic acid binding; INVOLVED IN: biological_process unknown; EXPRESSED IN: 11 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal protein L9/RNase H1, N-terminal (InterPro:IPR009027), Polynucleotidyl transferase, ribonuclease H fold (InterPro:IPR012337), Ribonuclease H (InterPro:IPR002156); BEST Arabidopsis thaliana protein match is: RNase H family protein (TAIR:AT5G51080.2); Has 3412 Blast hits to 3412 proteins in 738 species: Archae - 71; Bacteria - 1533; Metazoa - 7; Fungi - 2; Plants - 1270; Viruses - 0; Other Eukaryotes - 529 (source: NCBI BLink).	MQQVFKLCRQSSIPSPPFSQPAKLAFTIFHAAKAL AVATSEAANNGDSTSSSANDAVSWISLDVSKKH SRSLFNAAYVPLILSENDAVVAAETGSGKPHHARI NSGFGAHVHLSQKWSINQLVVKSISLSSSTSVFL CLGVVQWYSSKSNSSGKQKLKPKAVMEEEKGA YFVVRKGDIVGVYKSLSECQAQVGSSVSDLSATV YKGNSLPKDAEEYLASRGLHNALYTIKASDLKDC PFGKIVPCPFEKTCILHFDGSSKGNPIRAGAGAVL RTLNGNLICRIRQGLGNATSSVAEYRAFILGLKKA LEMGCTRIQAQGDNKIVCMQGSYVLLNPSYGND RLSNIYPI
355	sp_070600_cqhx	0.933333	#N/A	#N/A	Xanthine/uracil permease family protein; FUNCTIONS IN: transmembrane transporter activity; INVOLVED IN: transport, transmembrane transport; LOCATED IN: membrane; CONTAINS InterPro DOMAIN/s: Xanthine/uracil/vitamin C permease (InterPro:IPR006043); BEST Arabidopsis thaliana protein match is: Xanthine/uracil/vitamin C permease (TAIR:AT5G25420.1); Has 8520 Blast hits to 8482 proteins in 1844 species: Archae - 65; Bacteria - 6911; Metazoa - 346; Fungi - 119; Plants - 452; Viruses - 1; Other Eukaryotes - 626 (source: NCBI BLink).	MSGGGNVGCGNAGGGNNTVAVQEVKPHPVME TLSGIQYCINTNPPLFESIILGFQHSFLTLGIITVISS LTVPQMGGSNHDKAKVIQSMMFVSGTSTLLQTL VGTRLPVAIGASHAFVIPMTTIVLTTRLSCINQPQF LDQTFEKTMAEMQGALLVSSVVQIILGFSGIWRIF VRYLSPLSMVPLITFTGLGLYSLGFPMLADCVEIG LPHLILLLIFSQYMSGFLKTSRLIFDRYAIIFTTAIIWI YALILTASGVHNHRPPATQLSCRTDRAGLMLEAD WVYLPRPFQWGKPTVNMREATPIVFAGLVALIES TGTFIAAARYGSATPVPPSVISRGTFWLGIGTLLN AVFGAVAGPTASVESAGLLGLTRIGSRRVIQISAG FMLFFSIFGKFTAFFASVPFSIMAATYCILYAYVSS AGLNHLQFCNLNSFRTLFILGFSSFMAISVPQQFR DYLLTSNSGQLPTRSKWFFDTLATIFVSSPAVAIM VAILLDNTISEEEALRNDSGKRWWERFIQFTKDIR NAEFYSLPCELGNCFPSL
356	sp_070640_izrr	0.933333	#N/A	#N/A	terpene synthase 14 (TPS14); CONTAINS InterPro DOMAIN/s: Terpene synthase, metal-binding domain (InterPro:IPR005630), Terpenoid synthase (InterPro:IPR008949), Terpenoid cyclases/protein prenyltransferase alpha-alpha toroid (InterPro:IPR008930), Terpene synthase-like (InterPro:IPR001906); BEST Arabidopsis thaliana protein match is: terpene synthase 02 (TAIR:AT4G16730.1); Has 1701 Blast hits to 1676 proteins in 178 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1696; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MVVSIVFATTFNTTPIPTTKITVYFKESFINRHHSS NTKEYRLEHGVRKIRGIMTSKTRGYDIFNNLKDK DGQLKQELYKEDVKGLLSVCECSILGETNDISEQ LLRTFLNNMEEQSKASITIDNTLRYPYHKNLSILM VQNVLDEFDHVSFKFLHKSNNVKEWLNEVQELA RVDFGNARMTYRYEVQQISIWKNKHNFDKELSF ARIDPFKWWMWSYGILQGPNMSEDRIELAKAISF IYMIDDIFDGNGTSYDELLLFTEAINGWEYTDSIN QLPNCMKVCFKALLETTNDFSSKILTKHGWNPE QCLRKLVQVLA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
357	sp_071950_ikaf	0.993119	#N/A	#N/A	ATBFRUCT1; FUNCTIONS IN: hydrolase activity, hydrolyzing O-glycosyl compounds, beta-fructofuranosidase activity; INVOLVED IN: response to karrikin, response to wounding; LOCATED IN: cell wall; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 32 (InterPro:IPR001362), Glycoside hydrolase, family 32, active site (InterPro:IPR018053), Glycosyl hydrolases family 32, N-terminal (InterPro:IPR013148), Glycosyl hydrolase family 32, C-terminal (InterPro:IPR013189), Concanavalin A-like lectin/glucanase (InterPro:IPR008985); BEST Arabidopsis thaliana protein match is: beta-fructofuranosidase 5 (TAIR:AT1G55120.1); Has 4230 Blast hits to 4174 proteins in 1237 species: Archae - 18; Bacteria - 2583; Metazoa - 93; Fungi - 290; Plants - 1044; Viruses - 0; Other Eukaryotes - 202 (source: NCBI BLink).	MTSIRTVLGLAASLDLKLEQLDVKTAFLHGDLHEE IYMDQPGGFDEKGKDKLVFKLKKSLYGLKQAPR QWYRKFDSFMTSNGYKCTTADPCVYFRKFPGG NFILLLLYVDDMLIVGQDANVIQNLKKAFSKLFDMK DLGPSKKILGMEIARDRKVGKLWLSQEKYIEWVL ERFNMKHAKPVSTPLATHFKLSKRAEGNNALGP SMHLMV
358	sp_072040_guid	0.993334	#N/A	#N/A		MALCYSTKLVILLLGGSLLLGYGVIELQASHHVQN ATTLDQSLISADQPYRTAYHFQPVQNWMNDPCG PVIYKGIYHLFYQYNPIGNGSPDGPFVWGHSTSF DMVNWSPQPLILEPQMVSNMNGSYSGSITIIQGY IPAILFTGITPQNEQVQDLAFPKDISDPYLKEWVL VPENPVMSPTLENNINASSFRDPTTAWRLSDGN WRVLIGSKIDRLGLALLFRSMDFINWVQADEPLY SYDQTGMWECPDFFPVYINGFPIGVDTSAIGPTV KHVLKVSLDDTKRDIYAIGTYDTTKDAYIPDAGFL NDSSLRYDYGKYYASKTFYDNLKMQRILLSWINE SSTVEDDIKKGWSGIHAIPRRVWLDKSRKQLIQW PIAEIERLRVNPVILPTQVIKGGSTIEIPRITAAQAD VEISFRVRDMPIERFDASWTNPQLLLCSQKGPTV SGGLGPFGLLTLASTGLEEYTAVFFFRIFKAYDTNK FIVLMCSDQSRSSLNPTPDKTTYGTFVNVDPIRE GLSLRVLIDHSIVESFANKGKNVITARVYPTLAIYD KTRLYAFNYGVNNVDITWLTAWSMRKANIA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
359	sp_072870_funr	0.844039	#N/A	#N/A	UDP-glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: response to salt stress, N-terminal protein myristoylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: guard cell; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-Glycosyltransferase superfamily protein (TAIR:AT4G27570.1); Has 5043 Blast hits to 5008 proteins in 241 species: Archae - 0; Bacteria - 56; Metazoa - 380; Fungi - 15; Plants - 4575; Viruses - 3; Other Eukaryotes - 14 (source: NCBI BLink).	MDRTRPEVESILSTQKPDLVFYDMAHWIPDIASK LGIKSVSYNSVCAISMSHVRPSVLPKAAATSQVA LSSSSYVGSVPQWSRNDHNSSVYFGDEITLLER SIISLSSADAIAIRTCREIEGVYCDRVAATFNKPVL VTSHALPDLELQLSPLETRWADWLARFGPRSVIF CCLGSQHVLDAAQLQELALGLEMTGLPFLMAVK PPVGYASLKELLPEGYNDRVSGRGVVHGGWVQ CQQILAHPSLGCFVTHCGSSSMWEGLVSETQLV LLPQLPDQTLNAKLMADELKVGVKVERQQNGWV SKQGLCEAINRVMDEDSDISHVVRKNHAKYTSML ISPDFITGYIDNFIKDLQALIS
360	sp_073010_xeas	0.823119	#N/A	#N/A	uncharacterized protein	KQKRSNQQVLKPSSARGPRFSATAPLVDSPLLA DSSFLTDSSFLADSPLQEPQEESGTLDSIGASLG RLRKKDPSLTAEMIAALWCLHFLQTTLFMGHGYI
361	sp_073891z2584	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
362	sp_074000_sfue	0.823119	#N/A	#N/A	uncharacterized protein	MLFKEEKIRMCNWTSMVVLNFWYHGKFIFHPKR GYEQGLSETIISSRNRLSLTHMHKLIEKLGCKKEY TIWYVEDMKRFHEDGIAIKDPDDFYNLVELTKSNL VIHIVVQHRDVEEEEGKGLQNERWRLTQQRLPK NLG
363	sp_074680_doaf	0.926276	#N/A	#N/A	Concanavalin A-like lectin protein kinase family protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Legume lectin, beta chain (InterPro:IPR001220), Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Concanavalin A-like lectin/glucanase, subgroup (InterPro:IPR013320), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Concanavalin A-like lectin/glucanase (InterPro:IPR008985); BEST Arabidopsis thaliana protein match is: Concanavalin A-like lectin protein kinase family protein (TAIR:AT5G65600.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MLVVLLVFTDHVQSLNFSFPSFNPNTPNITYRND STPTNGAIQLTINRAVGSLSGSVGRASYNHPVRL WDQSSNRTTNFTTHFSFDIIQVIPDPNYYGDGIAF FIAPFDDSANAPADSFGGCLGLIGRGQDESCNST TEYPFVAVEFDTFHNAWDPGESGAPSNDHVGIN VNSVFSRTNFRLRESLKNRTTGNAWITYDSSTNN LSVFLSYDANPVFNGESNLSYVIDLSSILPERVRV GFSSSTGANYELHNILSWNFNSTLEDRDIILPAPR QSNNKAALVGGIVAGVSALLILLGLVAFMWWRR KRRDDMKFDVDDDFDEETGPRRFTYKELSQATN NFNEEGKLGEGGFGGVYRGFMTVMSREIAVKKI SRGSKQGKKEYVSEVKAISRLRHKNLVQLLGWC HEQGHLLLVYEYMPNGSLDSHLYTGKRKVVLPW DERYRIAQGLASAVLYLHEEWEQCVLHRDIKSSN VMLDSNFNAKLGDFGLARLVDHNLGSQTTEEAS KVSMIEWVWRLYGEGNLLEAVDKTIQGEFDIQQI ECLMVVGLWCCHPEYTFRPSIRQVINVLNLESPL PNLPMKFPVPVYCSPPTNVKNFYISSSSSSSSSSY SYSSSGFTGSSTGTTSSSSQSVVSGTAPLLSQN LTV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
364	sp_076670_xjri	0.8268478	#N/A	#N/A	cellulose synthase like G2 (CSLG2); FUNCTIONS IN: cellulose synthase activity, transferase activity, transferring glycosyl groups, transferase activity; INVOLVED IN: cellulose biosynthetic process, polysaccharide biosynthetic process; LOCATED IN: membrane; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Cellulose synthase (InterPro:IPR005150); BEST Arabidopsis thaliana protein match is: cellulose synthase like G1 (TAIR:AT4G24010.1); Has 2950 Blast hits to 2333 proteins in 442 species: Archae - 3; Bacteria - 792; Metazoa - 2; Fungi - 20; Plants - 2101; Viruses - 0; Other Eukaryotes - 32 (source: NCBI BLink).	MENPFHEFHVHKLRATLNITYSLFYFLALTFIFYYR SAAFFLKKATGVLFAAELVFSFLWAIKQPMYWRP ITRTVFPEKLPKDEKLPGIDVLICTADPNKEPTFEV INTVISAMSLDYPAHKLSVYLSDDAGASVTLNAM KQAWVFASWWLPFCRKYHLKTICPQAYFQNPE DQEESSCQTNDMDFVQHSMLLEEKYEAFKQQV RSNKLSGDVDNNNTSRDHPPLVQVINENCIEDG AVTLNPEAGEMPLLVYVSREKRPSHPHCFKAGA LNALLRVSSIMSNSPYTLILDCDMYCNDSISARQA MCFFIDPKISSSLGWVQFPQKFHNISDTDIYDSQL RVTWPVYWPGADGLQGPCIIGTNVYIKRKVLYSI NITTDDGINLKEVRSSLGPSNELTNSVVQNDKLN KTVKDKEVSSVMVQEAHFLASCTYEKGTMWGQ KVGFCYGTIVEDAKTSIRLQNKGWRSIYLNPTRP QFLGSATTSLNEMLVQGIRWYGGLLDLCLSKYC PLIYRPSRMQFLQKMYYSWMVLLALDFIPVLCFAI IPPICFLYGIPVYPKLSDPGFVAFAFVFVSTQLKHL SDVVVYSGGSVSAWVNEQRTWMAKCLSCYLYG TIVCIITKLGLHEATFTLTDKLQGDTTNWYDKGIYD FRTSKMFLVPIVTAVILNLCCFIGGVARLIVIRNWD AMFGQMLFSLYVLVMGFPVLEGVLLRKDNARISV SASMVSTILSLLLLSLGYLILRH	
365	sp_078000_hnxy	0.8577125	#N/A	#N/A	Protein kinase superfamily protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: chloroplast; EXPRESSED IN: shoot apex, embryo, root, flower, seed; EXPRESSED DURING: F mature embryo stage, petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: D6 protein kinase like 2 (TAIR:AT5G47750.1); Has 111692 Blast hits to 81240 proteins in 2732 species: Archae - 30; Bacteria - 14475; Metazoa - 44429; Fungi - 12685; Plants - 18495; Viruses - 325; Other Eukaryotes - 21253 (source: NCBI BLink).	MNKPRKQRSDLETESMNSLFQDLTVKDSVSVSL SVSLSPCSSTTASGSDVGGSSKPPNIGCLLKLVE NGESEKSASSSLSNSSNGVDNTNDASSSFRSSC PNKPHKGNDIRWEAIQCVKGGKDGELGLSHFRL LKKLGFGDIGSVYLAELRGMGCLFAMKVMDKGM LAGRKKLQRAQTEKDILGLLDHPFLPTLYSHFETD KFSCLLMEFCNGGDLHILRQRQPNKHFSEQASR FYASEVLLALEYLHMMGVVYRDLKPENVLVRED GHIMLSDFDLSLRCCVNPTLVKGSYQPNCTNSSY CIEPRCIEPACKLPVCVEPSCLQPSCFRPRLFNSK SSKVKREKFNFTNTSSLPILIAEPTNARSMSFVGT HEYLAPEIIRGDGHGSAVDWWTFGIFLYELLHGK TPFKGNGNRETLFNVVGQSLKFPEGSPASFAAK DLIRGLLVKDPQKRLGFKRGATEIKQHPFFESVN WALIRGSTPPEIPKPTDLTIYNHTSKTSSEPPSEK QGASDSDRSSGPYLDFEEFF	

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
366	sp_078820_rlkx	0.923119	#N/A	#N/A	Aminotransferase-like, plant mobile domain family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aminotransferase-like, plant mobile domain (InterPro:IPR019557); BEST Arabidopsis thaliana protein match is: Aminotransferase-like, plant mobile domain family protein (TAIR:AT2G25010.1); Has 691 Blast hits to 683 proteins in 18 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 691; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	TFLSGPTDNFDPRWIGMVEDVSTLGDYCWGDL GYATLVGQMSLAVRDSDPWRRHFLITLAGVPRLI ELWAFEHLPWLAPRKGQRPLEYPAGRRWGWK KKLTARPPPDTVWDLIRDGNPEHVVWTPWLSFR GTYASVRDSYALGQMRYLFVGRRDPVWYLGER
367	sp_079440_ptti	0.923119	#N/A	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G05858.1); Has 48 Blast hits to 48 proteins in 8 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGGSSSKSKRASTRTNHLDEPTPRRKVFPSDED RGYWVAEPGIDRKASAFIAKFHEARRFEAQTMAI PNSQTS
368	sp_080820_styp	0.923119	#N/A	#N/A	uncharacterized protein	YVAAADCSMLLLSAVCCCSVQYAAAQCSMLLLT AVCSTR
369	sp_080861z2824	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
370	sp_081120_kzde	0.923119	#N/A	#N/A	uncharacterized protein	MCKCRSHGTTPLFGLQSSHLNICYYHQDLHQRP LRPGSRHRFCSDRCALLLIGAWHLPRRPGIGHAL QRHPFSGLVDSAGELLHTP
371	sp_081530_scku	0.927137	#N/A	#N/A	fatty acid reductase 1 (FAR1); FUNCTIONS IN: oxidoreductase activity, acting on the CH-CH group of donors, fatty-acyl-CoA reductase (alcohol-forming) activity, long-chain-fatty-acyl-CoA reductase activity; INVOLVED IN: response to salt stress, microsporogenesis, response to wounding, suberin biosynthetic process; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Male sterility (InterPro:IPR004262), NAD(P)-binding domain (InterPro:IPR016040), Male sterility, NAD-binding (InterPro:IPR013120); BEST Arabidopsis thaliana protein match is: fatty acid reductase 4 (TAIR:AT3G44540.1); Has 2592 Blast hits to 2555 proteins in 497 species: Archae - 6; Bacteria - 759; Metazoa - 1002; Fungi - 307; Plants - 285; Viruses - 0; Other Eukaryotes - 233 (source: NCBI BLink).	MDQDSILKFLDNKAILVTGGAGFLAKIFVEKVLRS QPNVNKVYLLLRAPDNNSASSRLQNEVRTYYIFA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
372	sp_082770_zfhh	0.923923	#N/A	#N/A	N-MYC downregulated-like 2 (NDL2); CONTAINS InterPro DOMAIN/s: Pollen specific protein SF21 (InterPro:IPR015511), Ndr (InterPro:IPR000442); BEST Arabidopsis thaliana protein match is: N-MYC downregulated-like 1 (TAIR:AT5G56750.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGDSSDSISVHESISDSKEYYVKTGSGPVSVAV YGDQDKPALVTYPDLALNYASCFQGLFLCPEALS LLVHNFCIYHISPPGHEWGAAPISYEYEASTPSVD DLADQIVEVLDYFGLGEVMCMGVTAGAYILTLFA TKYKQRVLGLILISPLCRAPSWTEWLYNKVMSNL LYFYGMCSVVKEMLLKRYFSKELRGNSLVPESDI VQACRKLLDDRESTNVMHFLEAINRRTDISEGLK KLRCRSLICVGEDSSFHSEAVHMATKLDRRFSAL VEVQACGSLVTEEQPQAMLIPLEYFLMGYGFYR QSYKSLSPRSPLSPSCISPELLSPESMGLKLKPIK TRISLEV
373	sp_083530_gyrt	0.9355639	#N/A	#N/A	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein; FUNCTIONS IN: pyridoxal phosphate binding, carboxy-lyase activity, catalytic activity, tyrosine decarboxylase activity; INVOLVED IN: response to wounding, cellular amino acid metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aromatic-L-amino-acid decarboxylase (InterPro:IPR010977), Pyridoxal phosphate-dependent transferase, major domain (InterPro:IPR015424), Pyridoxal phosphate-dependent decarboxylase (InterPro:IPR002129), Pyridoxal-phosphate binding site (InterPro:IPR021115), Pyridoxal phosphate-dependent transferase, major region, subdomain 1 (InterPro:IPR015421), Pyridoxal phosphate-dependent transferase, major region, subdomain 2 (InterPro:IPR015422); BEST Arabidopsis thaliana protein match is: L-tyrosine decarboxylase (TAIR:AT4G28680.1); Has 5452 Blast hits to 5431 proteins in 1792 species: Archae - 99; Bacteria - 1888; Metazoa - 2300; Fungi - 264; Plants - 262; Viruses - 5; Other Eukaryotes - 634 (source: NCBI BLink).	MGSLDIKSLESTTQMMNMNNNPLDPEEFRKQGY MVIDFLAEYYKSIEKYPVRSQVEAGYLRKKLPVSA PNDPVSMETILADIQTDIVPGLTHWQSPNYFAYF PSSGSTAGLLGETLASGFNVVGFNWISSPASTEL ESIVMDWLAEMLSLPKSFTFSGGGGGGVMMGTT CEAILTTITAARDSLLDKIGRHNINKLVVYGSDQTH CSFFKSAKIAGISPANFRTIKTTKANAFSLRGDSL RATIEADVNAGLIPLFVVATVGTTSTAAVDAVGPL CDVAKEYNMWVHVDAAYAGNGICPELRHMIDG VEKADSFSFNAHKWFLTTLDCCCLFVKDPNALVR CLSTNPEYLKNKATDTNQVVDYKDWQLTLSRRF RSLKMWMVIRSYGVDFLRNFLRSHVDMAIYFEK MVVKDTRFEIVVPRSFALVCFRLSASAVLECSGQ KFLEEETLNEINTELLESVNSDGKIYMTHSLVDGIF MIRFAVGATLTEIHHVTYAWKVVQEHATALLARL SMIAC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
374	sp_085330_zifo	0.9385379	#N/A	#N/A	cytochrome P450, family 710, subfamily A, polypeptide 1 (CYP710A1); FUNCTIONS IN: C-22 sterol desaturase activity, oxygen binding; INVOLVED IN: oxidation reduction; LOCATED IN: endomembrane system; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, conserved site (InterPro:IPR017972), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 710, subfamily A, polypeptide 2 (TAIR:AT2G34490.1); Has 25674 Blast hits to 25623 proteins in 1435 species: Archae - 46; Bacteria - 2777; Metazoa - 10394; Fungi - 4411; Plants - 7239; Viruses - 0; Other Eukaryotes - 807 (source: NCBI BLink).	MSYNMSDIWAILILAAPFVATLVALVILIEQITYLKK KSVAPGPTLVLPFLGNVTGLIRNPEKFWDDQADY AKSSPLGISVNYIIGKFXAKSSPLGISVNYIIGKFMV FTRDSEMSHKIFANVRPDAFQMVGHPFGKKLFG EHNMINLYGQEHKDLRRRIAPNFTPKALATYSDIQ QLVILKHLTRWVENNDPKVALRVLARDMNLETSQ TVFVGPYLDEEKREGFKVDYNYFNVGVLALPIDIP GFAFRNARLAVERLVERLTRCVEESKKKMVENE EPTCLVDYWMQGMVNEERELQQPLQFPPSPPP PPPPNCSNHQIGAHMFDFLFAAQDASTSALLWA VTLLDSHPEVLKKVREEVESIWKVESNKLITPEQL VSMKYTQAVAREVVRYRAPATLVPHMVHQDFQL AENFTIPKGSIVFPSVYESSFQGFSEPDRFDPDR FYLEERREDQVYKRNYLVFGSGAHQCVGQRYAL NLLVLFIAMFTSLIDFKRHRTEGCDDLNFCPTICPK DDCLVSLSKRCANFPALSLP
375	sp_086041z3059	0.9323119	#N/A	#N/A	uncharacterized protein	#N/A
376	sp_086530_hkee	0.9327577	#N/A	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; Has 20719 Blast hits to 6096 proteins in 607 species: Archae - 22; Bacteria - 3243; Metazoa - 4364; Fungi - 2270; Plants - 237; Viruses - 128; Other Eukaryotes - 10455 (source: NCBI BLink).	MAPLPLFFRFTFYFLLFASSLFASSYSHRIFLSRR NLLDTNSKKNQTKLIKPSSSSSSSKNQTISSSNSN SSKKNQTKLAKPISKTSSKNVGAPPNTVGKPKVN STKSSDPLKSKSITKNKTKSSPSSTPTTKLPSSDT KSDQNQKQKQKQKLKSPSPPTWMQSQDDDND TDYDFMAEFRDLSNKFHQTFIPDMHFFSTTSKMY ITKANKEITQSFIKPYVSKQYHSSASTAISFAFIVP FLLVSLLFNRITAYFSLQKLLIFIQIYLSIYFSILCFSS LITGLEPLRFFYATSQSTYVW'LQVLQTLGYVMYL LMLLMYLILVFSTESGLGQKFLGLAQTFVGFAVG FHYYMTVFHRVVLHLPPKTNWKVHGIYATCFLVI CLLARADRTKKAYHDEEGVEEGKIN
377	sp_087591z3096	0.9323119	#N/A	#N/A	uncharacterized protein	#N/A
378	sp_087890_qxaw	0.9323946	#N/A	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G56260.2); Has 477 Blast hits to 219 proteins in 41 species: Archae - 0; Bacteria - 4; Metazoa - 91; Fungi - 61; Plants - 144; Viruses - 0; Other Eukaryotes - 177 (source: NCBI BLink).	MYQIKEDDKLFSKLVTKESSMANPSFRVYYGGN SVGVPFVWEAKPGTPKHTKHIDYSKEEEENKQY LPPLTPPPCYYLKQPNKLPTNGNSKRSSSRIFLG NLFLFPKSMHRRQISTTSSSSSSTLSCPSSSMLV SDASTMRSSDESRNLSLSLSLSLSLSRSGSGSGSG SGSGTRRRRTLSSPGSLFDFLVNGDQEDVEEED MRSPRSVFCFGLGCSRRIRSC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
379	sp_088320_effk	0.936338	#N/A	#N/A	MUTS-homologue 5 (MSH5); FUNCTIONS IN: damaged DNA binding, mismatched DNA binding, ATP binding; INVOLVED IN: mismatch repair, chiasma assembly, reciprocal meiotic recombination; LOCATED IN: condensed nuclear chromosome; EXPRESSED IN: male gametophyte, flower, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, IL.00 inflorescence just visible; CONTAINS InterPro DOMAIN/s: DNA mismatch repair protein MutS, core (InterPro:IPR007696), DNA mismatch repair protein MutS, C-terminal (InterPro:IPR000432); BEST Arabidopsis thaliana protein match is: MUTS homolog 2 (TAIR:AT3G18524.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGRRLLRNWFLRPILDAEILNSRLDAISFFLCSEE LMVSVRGTLKSIKDVPYILKKFNSPSSVSTTSDW AAFLKSLSSLLHLNKIFEVGISENLRGQANHFCLDI VDKAASCITTELSYIYELVIGIIDVTREKERGYGTIV KEGFCEELDELRQIYEELPELLEEVSSLELARLPN LHHGECVPCIVYIHQIGYLLWKFEEKLDECTLQNL EDLEFAFSDGENEPKRFFYRTSKTRELDNLLGDI YHKILDMERVITRDLISHVLLFSTHLLKAMNFAAEL DCLLSLALVAQQNNYVRPNLTTDAILDIQNGRHVL QEMTVDTFIPNDTKILDEGKINIITGPNYSGKSIYIK QVALIVFLSHIGSFVPADAAVVGLTDRIFCAMGSK LMAAEQSTFMIDLQQVGMMLRQATSRSLCLLDE FGKGTLTEDGIGLLGGTVNYFAGNDNTPKVFVCT HLTEIFHENLLPKSEKINFYTMSVLRPEESSMPAE DIVFLYRLVPGQAHLSYGLHCAMLAGVADNVVKR AAEILDATGKGNHVDRLSNENIQAQDQQYQKALE KLLAFDAFNGNLDLLFEDMFCTS	
380	sp_089871z3160	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
381	sp_090991z3216	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
382	sp_092191z3260	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
383	sp_092771z3313	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
384	sp_094840_irsr	0.949035	#N/A	#N/A	Protein of unknown function, DUF584; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF584 (InterPro:IPR007608); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF584 (TAIR:AT4G21970.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MENNRYNWRQYRGEEFQEEDVWAVIKENGETT NNHLRNPRSISRRLPSAARMIPGVQKHQHIDDNS NNESELVQSSAPVNIPDWSKIYGKSPVKKMIYGG CSSSFDYEDDNDYDLVLEDDNEEYAGDGRLPPH EWLAKKLARSQISSFSVCEGVGRTLKGRDLSRV RNAILSKTGFL
385	sp_096961z3475	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
386	sp_098111z3526	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
387	sp_099160_fatq	0.923119	#N/A	#N/A	Peroxidase superfamily protein; FUNCTIONS IN: protein binding, peroxidase activity; INVOLVED IN: response to oxidative stress, oxidation reduction; LOCATED IN: apoplast, cell wall; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, C globular stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Haem peroxidase (InterPro:IPR010255), Plant peroxidase (InterPro:IPR000823), Peroxidases heam-ligand binding site (InterPro:IPR019793), Haem peroxidase, plant/fungal/bacterial (InterPro:IPR002016), Peroxidase, active site (InterPro:IPR019794); BEST Arabidopsis thaliana protein match is: Peroxidase superfamily protein (TAIR:AT5G58400.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	GCDGSILLNDTANFTGEKTAIPNVNSVRGYDVIDT IKSQVENICPGIVSCADIVAVAARDSVVALGGPS WTVPLGRRDSTTASLTLANSDIPAPFLNITELLRS FSNKGFSPQEMVALSGAHTIGQARCTTFPTRIYN ETNINSTFATSLQANCPRNGGDDNLSPLDRTSST TFDNAYFTNLLSLEGVLHTDQELFNGNGTGTTDS FVSSYSSSSSTFFTDFSDAMVKMGNVPGVELEE PCLVMLPC
388	sp_099861z3606	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
389	sp_101740_azug	0.923119	#N/A	#N/A	uncharacterized protein	MQCSPARILMLVKEGMDSWTKRANLILADSESN PAERGLLQNPKMAWAIGKPLTTGHVTLLVDAAW KGIQKRKNSEWGAGVGWQGERTTERHIRDSMRI CATSALQAECIEIWEGMKSVAGMSKNILIKTNCQ VAIKALSNPKFTDNHIRHVLADIIALASTLDVYSCV KVDRLTISPAHILATNARKGMTV
390	sp_103201z3740	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
391	sp_104821z3793	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
392	sp_105691z3827	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
393	sp_106341z3848	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
394	sp_106671z3860	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
395	sp_109480_ctfd	0.955838	#N/A	#N/A	unknown protein; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 10 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G42430.2).	MASTYLLHPTTISIPFKQLNLQSLQPRFFLNTQTR TIYVRDFRIRASDGVENADSYLNMWKNAVNREK KEIEFNAIAQKLPPIDGGVDEETPDDLEKKSFEFQ KILEVSKEERDKVQRMQVIDRAAAAIAAARALLKG RKADDVANSLSSESVRVRSDNSGVLSPETENGG IDSGSSYKRQSRISAVGPPGPDFWSWTPPPEDD LNRSDVSKIQKAPKTNPHPTNTVIEKERSADYLSI PLESTYSENNLIQNLPPFQSLMEVHKVEALNSSQ EATVMQEEIGSQFSAHAAEANHALHNVEEKPFY GINPEGSRVWKETGVETRTDGVVCRWTLNRGV TADGVVEWEEKYWEAADGLDYKELGSEKSGRD SMGNVWREYWKETMWQEHGLLHMEKTADKW GKNAKGEEWQEKWLEKYDASGKADKWAHKWC SIDPNTQVEIGHAHVWHERWGETYDRQGGSIKY TDKWAERCEGDDWVKWGDKWDENFDQNGHG VKQGETWWEGKYGDQWNRTWGERHNGSGWL HKYGKSSSGEHWDTHAQEDTWYERFPHFGFMH CFENSPQLREVKKPSDI
396	sp_110171z4101	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
397	sp_110320_uimq	0.923119	#N/A	#N/A	uncharacterized protein	MLAQSPKVPEVLVPGLAARFRIDNQHNCQSQGA HLSVGAKSHIVTVENKAEHQNVTSNVESEAVVG CTTKLLMGRLENEGQH
398	sp_111040_mkxq	0.923119	#N/A	#N/A	KOBITO (KOB1); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G57200.1); Has 94 Blast hits to 94 proteins in 32 species: Archae - 0; Bacteria - 14; Metazoa - 0; Fungi - 0; Plants - 75; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MDPLTHLSLDRSLICSSPAQAKSNTSSSDELTVN ASHVTPHFGSFMFNSNNSVLTPKVCITSTTAAGL EDILSWIFFHKVVGISNFFLFVEGKAATPNVTIVLE SLPGVKVIPRTIQLEKQQAKSRIWNETWLSGAFH KPCNGALFVKQTLNMEMAIVMAREAGMDWIFHV DTDELLHPAGTRNYTITTELLAKVPTDVDAIVLTNY ESAVERDDIKKPFVEVTLFKKNFGHLPNDTNFKL YKMATRGNPNYFLTYGNGKSAARIQDHLRPNGA HRWYNYVKEPKEIQLDEAAILHYTYARFSDLTSR RDRCKCQPTQEDVKRCFWLEFDRSAFIVASTET QEGMLRW
399	sp_111070_gdwh	0.947734	#N/A	#N/A	uncharacterized protein	MNDMMHAAPVLGTTSSEEKQLAVGYKLMLEEM AAIRKAKQDKYESDGQNQKIKTMNSLPNNILNVF ELSREESLQCVGQGGSYCNNVFGPDCCSTSFACI PPGIVGGVCVA
400	sp_113330_kecn	0.933387	#N/A	#N/A	uncharacterized protein	MVVSVSTNNQTQKTLFSFIFIFILFSTKTTLLPCQA RLLELNSKITTSSSLPCEGDINLCEANCETLNGCR FTCIDDVGFYCLCHCDGESERFVETPRDVSSLAA MR
401	sp_113751z4321	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
402	sp_114251z4332	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
403	sp_114381z4337	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
404	sp_114891z4370	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
405	sp_117601z4468	0.9268079	#N/A	#N/A	uncharacterized protein	#N/A
406	sp_117630_eqqm	0.9284486	#N/A	#N/A	NAC domain containing protein 41 (NAC041); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: multicellular organismal development, regulation of transcription; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: No apical meristem (NAM) protein (InterPro:IPR003441); BEST Arabidopsis thaliana protein match is: NAC domain containing protein 83 (TAIR:AT5G13180.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MKKHNFVIEGVSRLPPGFRFQPTDEELLFQYLRC KIFGLPLPASVIPEVHFSNLDPWELPGDSKEERY VFTKRGAKHKNGDRNHTRATKATNSGYWEETG SEEHMSPPPNQNLGMPIKGVKKTLVFNIGKSPNY TSPTDWFMHEYSLVYKNHQGTSKMEDWVIYHVF LKQRVEENDDAIMLDYDNNNGHYNLTSNIDRYPC CSSSSSSSSNSFAEVSSYDSQSDEESSYNFL
407	sp_121070_qtrw	0.9283119	#N/A	#N/A	uncharacterized protein	MMIKNNEIVCMAGVHPGEQVVDKPASVVDLVSK SKSGGDRAEQADPLDGIPANVRSQYLEEVARQV RLSGGMFYSNVVQTYQTSSGSSSYSPRHPKAIV FPIAQEDKQAALVEDEHIPNTPHFSTRERAAICRS IYKLFVRILFCKNRKKKKAVELKAQVSDVVHHGDY AFMSIEEVRIEMQTTIDLQAMEVAALKSSKAELLN KSLQQDKYMMVLLEENKKAAAEKVEFQAQLEDY PQVKELANHIFDARRNYFAACDDRWARPVF
408	sp_121851z4625	0.9283119	#N/A	#N/A	uncharacterized protein	#N/A
409	sp_123481z4690	0.9331886	#N/A	#N/A	uncharacterized protein	#N/A
410	sp_124020_rkzx	0.9283119	#N/A	#N/A	receptor like protein 2 (RLP2); CONTAINS InterPro DOMAIN/s: Leucine-rich repeat-containing N-terminal domain, type 2 (InterPro:IPR013210), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: Leucine-rich receptor-like protein kinase family protein (TAIR:AT1G72300.1); Has 90767 Blast hits to 27052 proteins in 1087 species: Archae - 40; Bacteria - 7835; Metazoa - 18686; Fungi - 1120; Plants - 55519; Viruses - 2; Other Eukaryotes - 7565 (source: NCBI BLink).	MQIRNVPALAQINSFHLFILLLLYSSTFSYLQACNP NDQQSLLAFSNSLSSTLSPLNWSSDSDCCSSWE GIGCDDNGRVTILWLPLRGLEGGISSSLGTLTRLS KLNLSCNHLTGFLPDGLFSLLKSLEIFDISSNFLYG NLSASFTSDGSFASTIQVIDVSSNKFHGGIEPYW FFQGSNLSSFNASNNDFTGHIPSAICVTCPLLRTL DFSENTFSGQFTVGLGGCLKLRVLRAGFNYLSG VLPGDIYRLSSLIELSLPVNNIGGTIGKRISRLANL RIINLYSNNFAGTVPQEIGKLLKLEQLHLHNNLNG SIPASLMNCTNLVKLILRVNFLQGKIPDSIGNLPSL FYLDLSMNLLRGEFPLQLSRLPALISDRVAQKVN RSYLQLPVFVSPDSASSHQYNQLAHLPPAIYLKG NHLRGCIPVEISQLQKLHVLDLSKNSFSGRISPKF ANLTNLEMLDLSQNNLTGKIPTSLQTLNFLSKFNV SYNNLEGAIPIGGQFDTFTESSYISNPGLCGRVLQ SSCSAQPVQRGGYQDTADKSQKHILLEGFITGFS IIFSVVLAIQAYYHRRVLFYSKKR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
411	sp_124180_hsmy	0.923119	#N/A	#N/A	uncharacterized protein	MRRRTTAELSITESFMAKAEDSGIMDPVGEMEE GSEESGYGVERESGLGGLGLEGIMGSVDSDEG GFETEVGVVGNRGVISGDNAGESYGSILGSAAIV ESLNVVVLCEAEPV
412	sp_124461z4713	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
413	sp_124570_mcro	0.923119	#N/A	#N/A	Fumarylacetoacetate (FAA) hydrolase family; FUNCTIONS IN: copper ion binding, catalytic activity; INVOLVED IN: metabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Fumarylacetoacetase, C-terminal-like (InterPro:IPR002529), Fumarylacetoacetase, C-terminal-related (InterPro:IPR011234); BEST Arabidopsis thaliana protein match is: Fumarylacetoacetate (FAA) hydrolase family (TAIR:AT3G16700.1); Has 10822 Blast hits to 10670 proteins in 1907 species: Archae - 249; Bacteria - 6730; Metazoa - 306; Fungi - 435; Plants - 75; Viruses - 0; Other Eukaryotes - 3027 (source: NCBI BLink).	ATGLPWLLGKGLDTFTSISEMIPVSAVPDPHNLE LWFDVDGESKQKGSTKDMIFEIPFLISHIRSFMTL MEGDVILTGTSLV
414	sp_126420_sikk	0.942311	#N/A	#N/A	Transmembrane amino acid transporter family protein; CONTAINS InterPro DOMAIN/s: Amino acid transporter, transmembrane (InterPro:IPR013057); BEST Arabidopsis thaliana protein match is: Transmembrane amino acid transporter family protein (TAIR:AT3G28960.1); Has 5045 Blast hits to 4985 proteins in 317 species: Archae - 15; Bacteria - 153; Metazoa - 1796; Fungi - 1055; Plants - 1275; Viruses - 9; Other Eukaryotes - 742 (source: NCBI BLink).	NGVLIVPYALAAGGWWSLGLLFTISAAATYTSLL VKRCMEVDPRIKSYTHIGEYAFGEIGKIIVSIILYTD LYMVTTGFLILEGDNLNNLFPKLSLGIPIDGKSSFII LVALVLLPIVLFDNLSILAYISATGVFASILILVSVV WVGVFDGVGFHHVEDSGFFKWKGIPTAISLYML CYSSHPVFPALYTSMQKKHDFSKVLLICFTLSTFI YVSMAVLGYLMFGSSVESQITLNLPTDKVSSQIAI YTTLITPIAKYALILKPIVISTEGWFSNEYQNSKCF KVTIRIVLLASQVVIALALPYFGYFMSLTGALLCST ASLTMPCLCYLKITSTNNSRPGFIQQLLIWCIVFLS IAILIFGTYTSLVDILAEITH
415	sp_126650_jjdj	0.920156	#N/A	#N/A	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT5G59540.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGTMSVDITKSFQPAYDRVKELKAFDDTKSGVK GIVDSGAQTVPRIFIRPTDELSEDLNHPSVKRQVP VISFNGIDGTKERDVIQEYSNHVKNLSDVILRLLSV ALGLKPEHLKETECGEGWTLTNHYYPACPAPEL TMGTSKHSDPSFLTILLQDQIGGLQVFNENQWV DIQPIPGAFVVNIGDMLQISNDKLKSVNHRVIANL VGPRISVGLFLRG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
416	sp_126880_azjn	0.938636	#N/A	#N/A	alpha carbonic anhydrase 7 (ACA7); FUNCTIONS IN: carbonate dehydratase activity, zinc ion binding; INVOLVED IN: one-carbon metabolic process; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Carbonic anhydrase, alpha-class, catalytic domain (InterPro:IPR001148), Carbonic anhydrase, CAH1-like (InterPro:IPR018340), Carbonic anhydrase, alpha-class, conserved site (InterPro:IPR018338); BEST Arabidopsis thaliana protein match is: alpha carbonic anhydrase 5 (TAIR:AT1G08065.1); Has 3434 Blast hits to 3405 proteins in 557 species: Archae - 0; Bacteria - 714; Metazoa - 2114; Fungi - 83; Plants - 330; Viruses - 6; Other Eukaryotes - 187 (source: NCBI BLink).	MEKSIILVFSRALSLLLVLSFATSTIAQEVEDESEF NYIRGDVRGPENWGNIKDEWEKCKTGKMQSPIF IFPGLPERIPESDRVIRSYQTGQTTLKNRGHDIKL EWIEGGGTIRINGVDYILRQCHWHSPSEHAINGR RYALEVHFVHQSLDDKIAVIGFVYDYGNPDPFLS SIENELKLLISNKNSSEEIMSGIVNPSDVRMEGSR YYRYMGSLTTPPCDEGVIWTVETKVGSVSLEQV KLLRDAVNDDARENARPLQPLNWRRVSLYDTRL RIEDNDTLITPYVSSM
417	sp_127681z4859	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
418	sp_128150_hffd	0.923119	#N/A	#N/A	uncharacterized protein	MNIKLQGE
419	sp_129111z4909	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
420	sp_129520_zhzz	0.938627	#N/A	#N/A	CRINKLY4 related 3 (CCR3); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Regulator of chromosome condensation/beta-lactamase-inhibitor protein II (InterPro:IPR009091), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: CRINKLY4 related 4 (TAIR:AT5G47850.1); Has 117697 Blast hits to 115463 proteins in 4296 species: Archae - 105; Bacteria - 13273; Metazoa - 42642; Fungi - 9844; Plants - 33569; Viruses - 607; Other Eukaryotes - 17657 (source: NCBI BLink).	MKKKASCFSPLERCQFFLVLLTLFFFITMVSGFGS SSTIAASHGSNTVCGILAGKYQGIQCFQQGKTFS VLPNISYEAISGGQDFFCGLSSGGFDLFCWDTIF PSFNFHGKRIYHNTRFPLTDLTVGDDQVCAVQAR TGNIRCWEREARLGLLTPTLPRVEETAFLALTSG KGFSCGIVKDERNVLCWGESGIGDFIEREFGNM SMLSLVAGKSHVCGIAFSGLLICKGINDSGQLNVP LERPFHFSELALGRTYSCALQRSNGSLICWGNR GNINTSVITTDQKVQNLSFQSIISGVDYICGLRTDD FTVICWGPGLSGKLNTSVSLPSMVPGRCVRESC QLCGVLPDSESLCSGDTVICRSCVIELPLPLKPKI KLEAGNVPEPIPSPTAISPSSYPISSPPSKKETLR PLWAYEVFGFVGILSGMFAFAYFIYLKRPGSSKD GKKVLDVARDKQHDSVGSSNTGDSAFIASNASS APQSRSSSITQLSSMTLGRARSWDYSSKHVEKS EVFTLFELAAATKNFSLENRIGRGSFGTVYKGKL SNIGCEVAIKREEICTRKERGNAFESELVLLTRVH HKHLVSLVGFCEEQNERILVYEYMSNGALHDHL HTKNTHKSNGSVNSWKIRINIALDAARGIEYLHNY AVPPIIHRDIKSSNILLDANWVGRVSDFGLSLKGP ENQEGSMSIKPVGTVGYIDPEYYVLKVLTTKSDV YGFGVVLLELLTGKKAVFKEGEKSGPTGVVEYA GPFIVSGEVVKVLDKRVEIPVMNESEAVELLAYT AMMCVNLEGKERPTMSDVVSNLERALTLCKGSS GTFSFPNFYSFS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
421	sp_131590_nhhw	0.964786	#N/A	#N/A	uncharacterized protein	MFCLVMLLSSMYACGYSEKSHPPTNPGTLGTLG TPGRDGKPGKDGKPGKDGAPGQRGTLGGRGTP GGRGTPGTSGGTPGTPGTPGTPGTPGTPGTPR PPNTPHTLSALVSTKTRSNIGGPGAPSVPGTPGA PATSGTPGTPGTCGGKPGSPGAPATPGTPGTP GTPGSRSTPGSPGIPGSPGIPGVPGTPGAPGTH CM
422	sp_133801z5111	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
423	sp_134850_udxp	0.945074	#N/A	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G01990.1); Has 125 Blast hits to 123 proteins in 14 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 125; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAYTTKCCSSLGCWDLHNNLGFLNADILSSSYD DMENNNNNNNNNNNAASSLMLQASDYDNYSS SFSNDLLSTGYLEDALFEFSVRFKRRRVMVSNSN SNSNSNNNTNDFGVNNYDNDHQFATHHHFDDLS CWSNLFANYDMQSEDQELSHATLLNHHGHMHC SLGEITEKPIMKEMYSEVVKIEETNTLLLEEAIARR TPKLMDSSSYSKHNNSFNSKSTSFAYNSDTNST DSSSISVVTTDEEEKRKNNNNNNNNNNKTTTTTT TTRKMVYPFGLVKPGGKEGDITLKDINRRILMPPT RPLKHPVGDFACRPAVSPNGPGLS
424	sp_135570_aqht	0.920715	#N/A	#N/A	Ankyrin repeat family protein; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT4G03460.1); Has 38634 Blast hits to 17459 proteins in 653 species: Archae - 49; Bacteria - 2374; Metazoa - 20314; Fungi - 3164; Plants - 3215; Viruses - 213; Other Eukaryotes - 9305 (source: NCBI BLink).	MARIKQPINKGGAALWIKFVPGEDMSTVITFLVEK WSHVTCWPDANGSTPLHQAASLNTPYNLQVIRT ILYHWPQSAEVCDASGKTILHLMINRMPSYQEAT NLLKFKEIYALRNYQDQQGNTPLHIAAINLDINMV RVLLESSTKLSIKNMEGNSAASLIQQHNLLQMLR KRNMTCEESEAVDRADITFLRERMSKSGMDFLL SQDSKGRNVFHILMQIKNGSRILHDDFVDFIQQVL GNFPSLICQTDLKGDTPIHVLVRNHPDTTC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
425	sp_136640_nneo	0.923119	#N/A	#N/A	Protein kinase family protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT3G53840.1); Has 114140 Blast hits to 113200 proteins in 4569 species: Archae - 121; Bacteria - 12456; Metazoa - 42740; Fungi - 9504; Plants - 32338; Viruses - 412; Other Eukaryotes - 16569 (source: NCBI BLink).	MTHTKLKYKLLIQLQLVVVVLLSVTQKAVATLWC PDCGKTPVPYPLSTSQSCGDQLYKVRCTTGKLW LDTLNGSSYIITTISPKIQQLMISSQEFYRNSCTRA DLRSGGIWLNQTLPFSITDSNTILKFNCSKSTVRL SSMDCTSSSPCYKYMNKRCIGLHCCSYTTGGSK TAYRVDIAEASCTAYESFVKLNPALPVRKWPQSG IEIQWETPQVPICTLPLDCHGLANSTCLPDRAVA GQKKRCFCNKGFHWDPINGTCQSLTFAACGTVII IVFLVLVDQHKKHIKREAHRKLVKQREDVLNSHN HGKAAKLFSGKEMKNATKNFSKENLLGVGGFGE VFKGNLDDGTLAAIKRAKLGNTLGMDQQVLNEVCI LCQVNHRSLVRLLGCCLDLDQPVLVYEYVPNGTL YEHLHGRYSNEWGTLTWQRRLLIAKQTADGLAY LHFSAVPPIFHRDIKSSNILLDEKLNAKVSDFGLSR LVDTDATHISTCAQGTLGYLDPEYYQNYQLTDRS DVYSFGVVLLELLTSKKAIDFNRDEDDVNLAVYVTK LADREKLFDAVDAGLTTQASNLELETMKALGLLA VACVDEQRKNRPSMREVTEELEYIISIIDKETKSLL
426	sp_138800_opgi	0.945936	#N/A	#N/A	Protein kinase superfamily protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT3G62220.1).	IEWELGKGSYGRVLWGALKWACSSNKTGCKHT TRYEFLAQVSMVSRLKHESFVKLLVLMSCYAAVL LFCSTRSVPADLHLFVV
427	sp_139610_wxoe	0.923119	#N/A	#N/A	uncharacterized protein	MARNGDESDEETPEKDKDYDALNEKDQGEPEDI ANDEYLRRIDEFEELADAAALLGVDANHPTEGSS SKKGKGRGRGSGKARGPTKGLGKTGPMFLQFD SLLRPSGKWRLKYGGQIAALHHYITASLHCCPAV HCTACSRTAVQCTVGQQCSAATGQQCSAATGQ QCSEAHVRKTDQVVQQPMH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
428	sp_140380_uzax	0.968303	#N/A	#N/A	nucleolar protein gar2-related; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G57780.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MRESEKKKVQKNNPPKPSVRTPRRENKSVQEN VRKPLNAKESHTKASRAKPPSNSITSDKNVGAEP SEVYENVVIHYVDDAHRSEDASQGPKGDEDDVL EDSSSEFGKKGDDSDYESLKGSTSSQDDIQGVD DLKIDGASNVSNKIPNKGSLKVSDTQEKSNSGVK HSHSLPREPMKSDGSKVVPKDSSSECSEGTEDK PIEVTNQEICSGSDDDTVCAKGNMNTEQERAALK QKAEIMENRIEKLEDELRVVASLEISLYSVIPEHGS SAHKVHTPARRLSRLYIHACKHWTRSKRATIARN TVSGLILVSKTCGNDVARLTFWLSNTVVLREIISQ AFGISGQKSPARFSALNGSTKGSEGNAYPGKLK GHSSGKQVNKISQLFEDWEETSTFTSALEKIESW IFSRVVESVWWQALTPNMQKPVEDLIRQKSIGKL LGPSLGDQQQGTFSIELWKNAFQGAFERLCPVR AAGHECGCLPVLARMVMEQCVRRLDVAMFNAIL RESAHQIPTDPVSDPIVDSRVLPIPAGDFSFGSGA QLKNSVGNWSRWFSDKLDMEEDSLPIDDTETYS DDHNITHDKTKCFFLLNELSDLLMLPKDMLMDRSI RYEVCPSVNLGVIKRILCYFTPDEFCPDAVPGEVL EALNAESFVERKVGGEAFGSFPYGAPVVSYFPP SSGDVSEKVAETGGKPGVERNMCIVQKKGYMS DEELDDIGSPLTSFIEQSSPRILQNGNAGLKSLEQ TAGCIVNSRYELLREVWSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
429	sp_140720_fsxj	0.923119	#N/A	#N/A	cation/H+ exchanger 20 (CHX20); CONTAINS InterPro DOMAIN/s: Cation/H+ exchanger (InterPro:IPR006153); BEST Arabidopsis thaliana protein match is: cation/H+ exchanger 18 (TAIR:AT5G41610.1); Has 10038 Blast hits to 10006 proteins in 2135 species: Archae - 380; Bacteria - 7898; Metazoa - 11; Fungi - 167; Plants - 678; Viruses - 0; Other Eukaryotes - 904 (source: NCBI BLink).	MVSFSLCSTPLQGTEEFRRQRKQPSPVQVPRIV NNTNANQRLLMDIKLMKGRVEFLKHANKKVNEE HEKMKYEKMSMIAEGRSKTEIKNDNVYILLSLCDV ISVVGGILLGPSALGRSHQYMDKVFPKWSTPILE SVASIGLLFFLFLVGLELDLGSIRRSGKEAFSIALA GISLPFIGGAAVAVILRRTIEGADHAGYAPFLVFM GVSLSITAFPVLARILAELKLLTTKVGETAMAAAAF NDVAAWILALALAVALAGTGGEGGHKVSPLVSLW VLLAGVGFVAFMFLAIKPLMGWMVKRCSAEQGV VDELYTITTLAGVLVAGFITDFIGIHSIFGAFIFGLIP KEGEFAGKLLERIEDFVSGLLLPLYFASSGLKTDV TKIQGAVAWGLLVLVIAIACFGKIFGTFVVAKLCR MKGREALALGILMNTKGLVELIVLNIGKEKKVLND EMFAILVLMALFTTFITTPLVMAIFNPAKHGSDIET QKRKLQPDTSIHKDDTSSTSKSKQQFRILACAHG PGNVSGLINLIESTRSSAVKSNLKLYIMHLVELTE RSSSIMMVHRLRRNGLPFIDPTRRRSRVEVHDRV ASAFETYAQLGHVMVRPVTTISALSTMHDDVRHV AEGKRVMMILPFHKMWRKDEESGEMKMVENLG HGWRGVNQRMLKNSPSSIAVFLDRGFGGGRVT QNESTSQRQLGVCVVFFGGPDDREALELGARM MDHPEIRVTVIRFIETQGMDANSVKLMPSPEKCS DSNYTFSVTAMNTTEEKAEDEVMVEEFRNKWE GTMEFIQKQANNIEEAVLTLGRNGEYDLIVVGKG RFPSTMVAELADRQAEHAELGPIGDILSSSNNGIV SSVLVVQQHTMVHVEEMPVSRVVDGEDNHNDA TNQV
430	sp_144011z5562	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
431	sp_148351z5782	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
432	sp_149161z5855	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
433	sp_150531z5927	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
434	sp_151031z5939	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
435	sp_154140_gatn	0.923117	#N/A	#N/A	Protein of unknown function (DUF803); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF803 (InterPro:IPR008521); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF803) (TAIR:AT1G34470.1); Has 1265 Blast hits to 1244 proteins in 219 species: Archae - 0; Bacteria - 92; Metazoa - 420; Fungi - 372; Plants - 263; Viruses - 0; Other Eukaryotes - 118 (source: NCBI BLink).	MASSSDNSDPDKGMSADNIKGLVMALSSSFFIGA SFIVKKKGLKKAGASGVRAGVGGYSYLYEPLWW IGMITMIVGEIANFAAYAFAPAILVTPLGALSIISAV LAHVILKEKLHIFGMLGCVLCVVGSTTIVLHAPPE REIESVIEVWNLATEPGFLFYAALIMTTVFVLIFHVI PEYGQTHIMVYIGVCSLVGSLSVMSVKALGIAVKL TLSGINQLIYPQTWAFSMIVAACIITQMNYLNKAL DTFNTAVSPIYYVMFTSFTILASVIMFKDWNRQN PTQIVTEMCGFVTILSGTFLLHKTKDLSDCVSSCL PVKLSKHIGENGFEDSEGIPLLRRQDSVRST

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
436	sp_155690_sqcd	0.935108	#N/A	#N/A	Glycosyl hydrolase superfamily protein; FUNCTIONS IN: cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: carbohydrate metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, family 5 (InterPro:IPR001547), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: Glycosyl hydrolase superfamily protein (TAIR:AT4G28320.1); Has 827 Blast hits to 815 proteins in 205 species: Archae - 18; Bacteria - 244; Metazoa - 38; Fungi - 187; Plants - 310; Viruses - 0; Other Eukaryotes - 30 (source: NCBI BLink).	MKKTKNNSEWREKYAYPILGIATISTILYLNFNDQ LNFEFPIIFWQKNMGFVGVNSTNFIIINGHDDHHQ DDQRVYVNGWNSYWLMEASVWDPSRSKVSKM LKRGAQMGLNVCRTWAFSDGPGPNSLQILPGVF NERVLKGLDYVIVEARKNGIRLIFSLVNNLNAFGG KAQYVKWAQDAGINISSSTDSFFADSTIKGYYKA YIKAIVARKNSLSGVIYSEEPAIFAWELMNEPRCD SDCSSILQAWITEMVAYIKSLDQKHLVTIGLEGFY GPKTAEKSKVNPGEWAASLGTDFMQNSAIDGID FASVHAYPDSWMPHANTTAKLDFLQMWLDSHIS DGDNILKKPILFTEVGYPINDKGGGISYGDTSLKT VYDSIFESAKKGQAGAGALIWQLLVEGMEEYGD RFSLVAWEQPSTYKMMVEQSCKLSNINTGKRKS AGELNGKDPCSGIKYP
437	sp_155810_mczy	0.933706	#N/A	#N/A	curculin-like (mannose-binding) lectin family protein / PAN domain-containing protein; FUNCTIONS IN: sugar binding, protein kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation, recognition of pollen; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Curculin-like (mannose-binding) lectin (InterPro:IPR001480), PAN-2 domain (InterPro:IPR013227), Apple-like (InterPro:IPR003609), Protein kinase, catalytic domain (InterPro:IPR000719), S-locus glycoprotein (InterPro:IPR000858), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: lectin protein kinase family protein (TAIR:AT1G34300.1); Has 27641 Blast hits to 27042 proteins in 566 species: Archae - 0; Bacteria - 139; Metazoa - 3969; Fungi - 187; Plants - 22616; Viruses - 64; Other Eukaryotes - 666 (source: NCBI BLink).	MVTRCIFVGTSATMLCQFLVSLLFSVAVSEISLGS KLTVSENQLWVSSNGDFALGFFTHSDNVNQYSL GIRFHSGSIPVDKQSVVWVAGAELTVSNQSYFQL MEEGGLVLFDPSNGGVVWTSNTSGKSVASASL QDDGNLVVLDGNKDMVWQSFDTPSDTLLPGQV LSVHKTLRALSGNLVSSHYSLYLNDSGQLQLRW ETGVLYWVRGSPSNLSLKASLSSNGSFQLLDDR SEAFWSVFAEDHNDNVKFRFLRLDVDGNLRIYS WDEDSVSWKSVWQAVENQCDVFATCGDQGIC SFANSSLSICKCPFGSAMESNSKCLAPYRQQSC SSRAMVEYKHTSLYGIYPPSDTIIITSLDFCQRLCE KDPLCTAVTYLNDGSARCQIKKTRFITGSSDPSLS SISFVKTCTDPFAVLPNKSPTAASPSLPSLLSSKKS NQSSRFPIRPVIDAALVSFCLFCMLQIAVGFWVC KRRRATVRKKASPPSHGLISAGLVRWTYGEIKEL TGNFKDQLGPHSYKSVLQNNQPAVVKRLESQIE ERRFRAMVSMIGSIHHKNLVKVEGYCCESGNRFL VYELLKNGSLEDFMHDPASCKKLSWKNRMLICL DVAKAISYLHTGCRDFVSHGKLKCGNVILDENLE AKVTEFGLQKLDGIETYAEKDVEDFGKMVLYLIC GRQEEVCEWSYNKWVSGYGWRVVDKRIEGEV DFEELERGLRIAFWCLQVDERMRPSMGEVVNVL EGTLTVDPPPPPFSDSRNLQEDILAESELV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
438	sp_157070_pyrh	0.936756	#N/A	#N/A	bidirectional amino acid transporter 1 (BAT1); FUNCTIONS IN: arginine transmembrane transporter activity, L-lysine transmembrane transporter activity, L-alanine transmembrane transporter activity, L-glutamate transmembrane transporter activity; INVOLVED IN: amino acid transport, transport, transmembrane transport; LOCATED IN: integral to membrane, membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Amino acid permease subfamily (InterPro:IPR004756), Amino acid/polyamine transporter I (InterPro:IPR002293), Amino acid permease domain (InterPro:IPR004841); Has 15465 Blast hits to 15422 proteins in 1894 species: Archae - 296; Bacteria - 11987; Metazoa - 438; Fungi - 2169; Plants - 254; Viruses - 0; Other Eukaryotes - 321 (source: NCBI BLink).	MDSIDVDRQPLLGSERSSISDDNRLQQLGYKQEL NRSLSAIANFAVTFSIISVLTGLTTMFNMGLAYGG PVTMVYGWFIVGIMTMMVGLSMAEICSAFPTSG GLYFWSAKLCGNRWGPFASWLTGWFNIVGQW AVTTSVDYSLAQMIQVIILLSTGGKNGGGYEASKY VIMGFHGGLLLLHAMLNSLHISWLSFFGQFAAAW NVAGVFVLMILVPLVAKEKASAEFVFTHFNTDNG EGIKSIVYIFVLGLLMSQYTILGYDASAHMTEETKN ADKNGPKGIISAIVISIIVGWGYIVGITFAVTDIPYLL SSENDAGGYAIAQVFYLAFKTRYGTGIGGILCLCI VAVAVFFCGMSSMTSNSRMAYAFSRDGAMPMS SVWQKVNKQEIPINAVWMSAFISFCMALTSLGSL VAFEAMVSIATIGLYISYALPIFFRVTIARKSFVPGP FNLGRYGVVIGWIAVTWVITISILFSLPVAYPINTE TLNYTPVAVGGLLILTVSYWVFSAKSWFKGPLKN VTV	
439	sp_160031z6573	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
440	sp_161020_ustu	0.923366	#N/A	#N/A	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT3G61400.1); Has 8339 Blast hits to 8299 proteins in 1003 species: Archae - 0; Bacteria - 1128; Metazoa - 122; Fungi - 867; Plants - 4923; Viruses - 0; Other Eukaryotes - 1299 (source: NCBI BLink).	MVSSSKKSSSEIEYDFIKGDLKAFDESKLGYKGLVD AGVTTLPRIFVNQEATAFSRESEVVYTTSDNPHV NIPVIDLSGVVEDGGDLEKRREVMKAVSEACEK WGFFQAVNHGVPTHVLDEMINGAKRFHELDVES KKEYYMVRGVNSGRKFSYTSNMVLYTGKVTNW RDTTIAGMEPMPDPQDFPETCREITIEYAKHMKK VALTFLELISEAMGVESNYLKDMDCCQEFLHLTH YYPPCPQPELAIGINQHADNDIVTLLLQDEIGGLQ VLHDNQWHDIPHIPGAVIINTGDLLQLISNDKYKS VIHRVQSKKVGPRISVPCFFRPKTNNPRILAPIKEL LSPENPAIYRPITTTGEYIAHYHSTGQDYGVKPAL DHFKLAN
441	sp_163000_irwc	0.953636	#N/A	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G34090.1); Has 73 Blast hits to 73 proteins in 21 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 65; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	KRGFYYFAWRQTWHFPPCNTFVKKQLRCYDFS VTVANSTIWVDQYIHRMLVEEDACSPKDRNRELL HVSANAGKLYREGSFIDSQLPNLDTYHLRKVGVF PDILERKINKHLEKGNRVSALITGEFYAKKQHFPG FGRPVAFNAHILMRVGRDLEAKDAARGALKSPW WTLGYKYQEVAEIAKWKDEQICYVKKKLTDEGK REDLSRGKELAQIALDQAAFLLDLASVEGNWDDF VDRISECYKVAGLHDMARFIAYKS
442	sp_163030_xjsc	0.923119	#N/A	#N/A	uncharacterized protein	MPGAKFWEKVNLAAPAPPRYKVQPGRPKANKR RLEAGEKQGKGKKAKGDQPTQAEHPKRPKKQN KCGKCGGLGHNRTTCKSTGPSEKPANKGKGGR PTSNTPWVAEQRRKKEERVHKAAVATAACAGSS SQPQSQVTQSQQQSQPHQH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
443	sp_164920_emoj	0.933119	#N/A	#N/A	uncharacterized protein	EEKRGEAGECEAGKQGTVVEYGGVSGWPSGA GVVVEQWWEWWWSVAVVVVNGGCSTVVP
444	sp_165380_awih	0.956089	#N/A	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: root; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G17350.1); Has 261 Blast hits to 261 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 261; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MFCERFPDRCDTTCDPASYLCRHFPECAPNNKA LRALSCCSYYPKATESLRMMLKHCDTYTSVHWK SVGGLGTSQPYDQVPEYGVRVDFEIPLTASCLR CQDRAKGGGTCGFDTSNQDFVCLCEERNVTTY CKDHQVERHGGHTKAVAGTVTVASIAGVGIGAG V
445	sp_166571z7270	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
446	sp_166761z7281	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
447	sp_166881z7285	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
448	sp_168211z7384	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
449	sp_170411z7481	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
450	sp_171160_fxje	0.933973	#N/A	#N/A	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; FUNCTIONS IN: lipid binding; INVOLVED IN: lipid transport; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Bifunctional inhibitor/plant lipid transfer protein/seed storage (InterPro:IPR016140), Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor (InterPro:IPR003612), Plant lipid transfer protein/hydrophobic protein, helical domain (InterPro:IPR013770); BEST Arabidopsis thaliana protein match is: Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (TAIR:AT1G48750.1); Has 308 Blast hits to 307 proteins in 31 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 308; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MATKVATLTLLAAVVVAMLLVETPATEAVTCSPV ELTPCAPVIMYNQAPSNACCEKMQEQKPCLCGY AKNPTLKSYINSPGAKRMASTCKVDVKC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
451	sp_171340_utjn	0.994873	#N/A	#N/A	Peroxidase superfamily protein; FUNCTIONS IN: peroxidase activity, heme binding; INVOLVED IN: response to oxidative stress, oxidation reduction; LOCATED IN: cell wall, vacuole, membrane, plant-type cell wall; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Haem peroxidase (InterPro:IPR010255), Plant peroxidase (InterPro:IPR000823), Peroxidases heam-ligand binding site (InterPro:IPR019793), Haem peroxidase, plant/fungal/bacterial (InterPro:IPR002016), Peroxidase, active site (InterPro:IPR019794); BEST Arabidopsis thaliana protein match is: Peroxidase superfamily protein (TAIR:AT2G18150.1); Has 4523 Blast hits to 4498 proteins in 279 species: Archae - 0; Bacteria - 4; Metazoa - 1; Fungi - 207; Plants - 4264; Viruses - 0; Other Eukaryotes - 47 (source: NCBI BLink).	MGSSKFLPFLQLVLCSLLLGISNGQSIVPVVPGLS YTFYDSNCPGLEIIVRSHLWEVFNNNITQAAGLLR LHFHDCFVQGCDGSVLLDGSTSGPSEQDAPPNL TLRAAAFTIIDDLRALVHNACGPIVSCADIVALAAR ESVFLPNGPYYQIPLGRRDSVNFATRPVTLANLP PPTFNTTSLINAFAAKNLTTTDLVALSGGHTIGRG HCASFANRLYPTQDSTLDQTFANNLKLTCPTNTT DNSTNLDIRSPNLFDNKYFVDLMNRQGLFTSDQD LYTDSRTRNIVTSFAVNQSLFFEKFVDGMLKMGQ LGVLTGTQGEIRTNCSATNVNNINIDLASV
452	sp_171440_eirq	0.993813	#N/A	#N/A	expansin-like A2 (EXLA2); INVOLVED IN: response to cyclopentenone, plant-type cell wall organization, unidimensional cell growth, plant-type cell wall loosening; LOCATED IN: endomembrane system, extracellular region; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Barwin-related endoglucanase (InterPro:IPR009009), Pollen allergen, N-terminal (InterPro:IPR014734), Rare lipoprotein A (InterPro:IPR005132), Major pollen allergen Lol pI (InterPro:IPR005795), Expansin/Lol pI (InterPro:IPR007118), Expansin 45, endoglucanase-like (InterPro:IPR007112), Pollen allergen/expansin, C-terminal (InterPro:IPR007117); BEST Arabidopsis thaliana protein match is: expansin-like A1 (TAIR:AT3G45970.1); Has 1877 Blast hits to 1874 proteins in 122 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1869; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MSILISLMLFFFFFAYASACDQCVLSKASVFTKA SSLSSGSCGYGSLALKFNGGHVAAGVPSLYKDG AGCGACFQIRCKNSGICNKAGTKVIMSDVNKSNN STDFVLSSRALRAMALHGKDQQILKLGVVDVEYK RIPCDYKGQNLAVRVEESSKNPGYLAIKVLYQGG QTEIVAIDIARVGSSNWIFMSRKYGAVWETSKYP GPGPLQLRFVVTSGFDGKNVWAENVLPANWKT GVTYDSKVQIDDIALEGCTPCADSHWN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
453	sp_172180_sdkj	0.9946339	#N/A	#N/A	SIGNAL PEPTIDE PEPTIDASE-LIKE 2 (SPPL2); FUNCTIONS IN: peptidase activity, aspartic-type endopeptidase activity; INVOLVED IN: proteolysis; LOCATED IN: endosome, plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protease-associated PA (InterPro:IPR003137), Peptidase A22, presenilin signal peptide (InterPro:IPR006639), Peptidase A22B, signal peptide peptidase (InterPro:IPR007369); BEST Arabidopsis thaliana protein match is: SIGNAL PEPTIDE PEPTIDASE-LIKE 4 (TAIR:AT1G01650.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	VYATEVGTVLLSCAFFYDIFWVFASKWFFHESVM IMVSWVVLFNSEGGCFSIPLQL
454	sp_173760_jwjm	0.9933119	#N/A	#N/A	multidrug resistance-associated protein 14 (MRP14); FUNCTIONS IN: ATPase activity, coupled to transmembrane movement of substances; INVOLVED IN: transport, transmembrane transport; LOCATED IN: plasma membrane; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: ATPase, AAA+ type, core (InterPro:IPR003593), ABC transporter-like (InterPro:IPR003439), ABC transporter integral membrane type 1 (InterPro:IPR017940), ABC transporter, transmembrane domain, type 1 (InterPro:IPR011527), ABC transporter, transmembrane domain (InterPro:IPR001140), ABC transporter, conserved site (InterPro:IPR017871); BEST Arabidopsis thaliana protein match is: multidrug resistance-associated protein 6 (TAIR:AT3G21250.2); Has 673798 Blast hits to 360971 proteins in 4075 species: Archae - 12279; Bacteria - 544381; Metazoa - 13093; Fungi - 8967; Plants - 6941; Viruses - 21; Other Eukaryotes - 88116 (source: NCBI BLink).	MRLNGITKSMVANHLAESLAGAITIRAFQEEERFC SKNFDLIDKNARSFFNNFAATEWLIQRIEILSAVVL ATAALCMVLLPNGTFSSGFIGMALSYGLGLNLGM VFSIQNQCILANQIISVERLYQYMDIPSEALEVIND SRPPQNWPSVGKVEIYNLQIRYRIDTPLVLKGINC VIEGGDKVGIVGRTGSGKTTLIGALFRLVEPVDG KIIVDGVDICSIGLNDLRSHFGIIPQDPTLFNGSVR FNLDPLCQHTDEEIWEV
455	sp_174161z7883	0.9923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
456	sp_176310_fnxx	0.956937	#N/A	#N/A	Polyketide cyclase/dehydrase and lipid transport superfamily protein; CONTAINS InterPro DOMAIN/s: Polyketide cyclase/dehydrase (InterPro:IPR019587); BEST Arabidopsis thaliana protein match is: Polyketide cyclase/dehydrase and lipid transport superfamily protein (TAIR:AT4G32870.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MESVSSHKKWQGKARVELAGVKAEEVWPFLAD FCGLNKWFPTLSTCIRVEGVEGQPGCVRYCAGF KTHIEHANENHVNWTKQKLLSIDPDNMTFSYSIID GNVGFSSYVSKVRVLLQVDGCVIEWDYEVEPVE GWKPKDLDEFIGKGLKVMAKRMEEALKGRAMN
457	sp_177510_rknn	0.923119	#N/A	#N/A	uncharacterized protein	MIECCGGSVVLEGASGLRGDGVWWDEGLRCCG AAWLMRRRGELKMLLCLVVHFKECGVG
458	sp_181790_jczu	0.932226	#N/A	#N/A	aldehyde dehydrogenase 2B7 (ALDH2B7); FUNCTIONS IN: 3-chloroallyl aldehyde dehydrogenase activity, aldehyde dehydrogenase (NAD) activity; INVOLVED IN: oxidation reduction, metabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: LP.06 six leaves visible, 4 anthesis, LP.10 ten leaves visible, 4 leaf senescence stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Aldehyce/histidinol dehydrogenase (InterPro:IPR016161), Aldehyde dehydrogenase (InterPro:IPR015590), Aldehyde dehydrogenase, N-terminal (InterPro:IPR016162), Aldehyde dehydrogenase, conserved site (InterPro:IPR016160); BEST Arabidopsis thaliana protein match is: aldehyde dehydrogenase 2B4 (TAIR:AT3G48000.1); Has 60835 Blast hits to 60447 proteins in 2994 species: Archae - 474; Bacteria - 34339; Metazoa - 2649; Fungi - 2135; Plants - 1675; Viruses - 0; Other Eukaryotes - 19563 (source: NCBI BLink).	MNVDKLAFTGSRSTGKLVLELASKSNLKPFTLEL GGNSPFIICENADIHQAMETTCFALSFQPGTMLL CWVLYLCRP
459	sp_182041z8496	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
460	sp_182310_umxi	0.923119	#N/A	#N/A	uncharacterized protein	EAYEKNVAECREKGINKDPNQIYFETVGGRKKGK VPGLGSGAALYYEPSCRRGGSSSSSYTPSIYSQL SARLEETHQQLSQKSIELETTKAEMIKALDDERLL RQRERFEEERKERELERLERQRDVAELKRRMESY ERMFSQCTGSFSQSQDPRDPSGGGKPLC
461	sp_182991z8566	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
462	sp_183491z8707	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
463	sp_183550_aazg	0.933134	#N/A	#N/A	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: Transcription factor, SBP-box (InterPro:IPR004333); BEST Arabidopsis thaliana protein match is: Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein (TAIR:AT5G50670.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MEWDSESCGWNGVSKLEIEDNISYHHHLSPLAA GNTTAAIRYARGTSSKTPVVIVAGKEQRFCQQCS RFQSLEEFDEVKRSYRKRLDGHNRRRRKRQPEA FYMPNCAAGFLSSHTADGMVQFCCPQSHTVNW PTMSQPPDLYLRNNQETIYEPLLNGVGQIGNRTV KVSADNSGFSLSLLSRYSSDICVESAMQPTTMSS STGQGQGQGSCTSLHLNNSIVQFSCSHGLEDME TSASLSSRTHNITLQSGHDEFRDPRSSRIFPFGW E
464	sp_183911z8741	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
465	sp_184101z8837	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
466	sp_185291z8971	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
467	sp_185301z8976	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
468	sp_186321z9043	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
469	sp_187700_qkkc	0.932179	#N/A	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G73940.1); Has 58 Blast hits to 58 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 58; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MVFRTAIRLMTSKKPKPKMKPIELKKGQQPEQTQT ITRAIFDILKEQGPLTIAQTWERLQEVGLKDLTSKT QMKTVMKWMKERQKLKMICNHVGSHKQFLYTT WFTKENIKPAKPGSYDSKE
470	sp_190361z9640	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
471	sp_190601z9675	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
472	sp_192281z9788	0.933119	#N/A	#N/A	uncharacterized protein	#N/A
473	sp_192751z9929	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
474	sp_193180_pccw	0.933119	#N/A	#N/A	uncharacterized protein	MNAVKAALDTNDVDEYDVSVLINVTGKNKRRRRG GSSKSIPQKVKGKSKAQAEPDIPYERISVIEKSIS MNQYTRKNRRTKNSEDALHLDEPLHLEESAQIGT NKSGVALXRPSRTNLMTHCTLIQVSNTTWCTTNT NTNTNTT
475	sp_193311z9971	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
476	sp_193621z9987	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
477	sp_194260_zfku	0.923119	#N/A	#N/A	Peroxidase superfamily protein; FUNCTIONS IN: peroxidase activity, heme binding; INVOLVED IN: oxidation reduction, response to oxidative stress; LOCATED IN: endomembrane system; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Haem peroxidase (InterPro:IPR010255), Plant peroxidase (InterPro:IPR000823), Peroxidases heam-ligand binding site (InterPro:IPR019793), Haem peroxidase, plant/fungal/bacterial (InterPro:IPR002016), Peroxidase, active site (InterPro:IPR019794); BEST Arabidopsis thaliana protein match is: Peroxidase superfamily protein (TAIR:AT1G49570.1); Has 4658 Blast hits to 4636 proteins in 311 species: Archae - 0; Bacteria - 0; Metazoa - 8; Fungi - 376; Plants - 4203; Viruses - 0; Other Eukaryotes - 71 (source: NCBI BLink).	MNSNKYFSALFVIITLGLISAVSCQDLSCNYYDKD CENFEEIVAKKVWEWFKKDKTIAASLIRLHFHDC AVRGCDASILLDYPGSEKWATSSKTLRGYEVIDD IKATLEKKCPNLVSCADILTAAARDATVLVKGPYW SNEYGRKDGLISVAQEADDLPNGREDVTTLIEFY QALGLNMLDMVVLSGAHTIGRSTCGAIKNRLSRY DENGKLHSTIMAPTYLNYLRRKCKNDTNYVDLDA RTPYTFDNEYYKNIEKNMGLLYSDQALYTDHRTS EYIATLANDDGGSFPYLFSSSMVKLGNILGDAQD DGEGEVRQHCSRLNSGY
478	sp_194421z10073	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
479	sp_194650_ftdk	0.923942	#N/A	#N/A	ATMAP4K ALPHA1; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: response to salt stress, hyperosmotic response, response to wounding; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT3G15220.1); Has 132200 Blast hits to 130116 proteins in 4312 species: Archae - 185; Bacteria - 15420; Metazoa - 49247; Fungi - 12380; Plants - 32481; Viruses - 583; Other Eukaryotes - 21904 (source: NCBI BLink).	MADVAAFIEAAGSRFGDLELIGRGSFGDVYKGFD KELNKLVAIKVIDLEESEDDIEDIQKEISVLSQCRS QYITEYYGSYLHQTKLWIIMEYMAGGSVADLLQS GPPLDEISIACILRDLLHAIDYLHAEGKIHRDIKAAN ILLTENGDVKVADFGVSAQLTKTISRRK
480	sp_194691z10097	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
481	sp_19544 1z10201	0.923119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
482	sp_197700_ghsd	0.988867	#N/A	#N/A	Major facilitator superfamily protein; FUNCTIONS IN: carbohydrate transmembrane transporter activity, sugar:hydrogen symporter activity; INVOLVED IN: transport, transmembrane transport; LOCATED IN: integral to membrane, membrane; EXPRESSED IN: leaf whorl, sepal, flower; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Sugar transporter, conserved site (InterPro:IPR005829), Major facilitator superfamily (InterPro:IPR020846), Sugar/inositol transporter (InterPro:IPR003663), General substrate transporter (InterPro:IPR005828), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT4G04750.1); Has 23023 Blast hits to 22536 proteins in 1742 species: Archae - 399; Bacteria - 9001; Metazoa - 4095; Fungi - 6023; Plants - 2554; Viruses - 0; Other Eukaryotes - 951 (source: NCBI BLink).	MPTNSDRFDGDWKKIKALISSKTVIQVGYTTPTQ TSIVKDLNLSEAEWIDPRLP
483	sp_198561z10674	0.988319	#N/A	#N/A	uncharacterized protein	#N/A
484	sp_198890_xshx	0.988232	#N/A	#N/A	Cornichon family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: intracellular signaling pathway; LOCATED IN: endomembrane system, membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cornichon (InterPro:IPR003377); BEST Arabidopsis thaliana protein match is: Cornichon family protein (TAIR:AT1G12390.1); Has 605 Blast hits to 605 proteins in 175 species: Archae - 0; Bacteria - 0; Metazoa - 316; Fungi - 159; Plants - 93; Viruses - 0; Other Eukaryotes - 37 (source: NCBI BLink).	MHWDLILWLISLISNIVLLVLLIYQLICLSDLEADYM NPYETAANINSLILPEFGLQAAFCALFLVTGHFFM FLVTLPVAIYHARLFARGEHLVDVTEIFRALSVEK KHRLIKLGLYLLFFFLVIFRLVIAIYNSLADEEEALH GFWVF
485	sp_199031z10717	0.988319	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
486	sp_199840_kgne	0.923119	#N/A	#N/A	NIMA-related kinase 2 (NEK2); FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: NIMA-related kinase 3 (TAIR:AT5G28290.1); Has 124376 Blast hits to 122221 proteins in 4182 species: Archae - 113; Bacteria - 14280; Metazoa - 45892; Fungi - 12262; Plants - 30619; Viruses - 487; Other Eukaryotes - 20723 (source: NCBI BLink).	LSIFVKKKVHPVIESYADYGLRSLAVATQVCASGP MSMLQLHETLTYLFCELDMTKYIFCYCLSVGSTE ENKRESTVAPLPAKYSGAFRGLIKRLLLRNNPKLR PSADELLRHPLLQPYVEKIHLHTRNPMCDAIDDN WDDTDSNCIDKATLPEAEDERKIDPSMSEAELDS FYTDDRIKPTLSDLERRLAKSKAQDPFRRVAATK QVPSKASNDVRNPRLSRAKTFGTSKRRPELSKN KQSVSFDASM
487	sp_200701z11025	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
488	sp_200741z11027	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
489	sp_200981z11050	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
490	sp_203671z11605	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
491	sp_203861z11635	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
492	sp_204020_zoaq	0.923119	#N/A	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: anchored to plasma membrane, plasma membrane, anchored to membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G30700.1); Has 65 Blast hits to 65 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 65; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	ACPFNFIAAESIMGITATDCFTAFAPYLANVVCCP QLHATLVILIGQSSKKTMLFALNETLSKQCLSDVN QVLASQGANDNLQQICKVWESNLIEGSCLVKDV REFEKTIDTSKLMASC
493	sp_204651z11817	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
494	sp_204990_emnw	0.923119	#N/A	#N/A	uncharacterized protein	MKPKRLVFSCSRPPSSSFQSPPFPKLDKAERVIS SSTYQHW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
495	sp_206970_etzs	0.928079	#N/A	#N/A	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: pentatricopeptide (PPR) repeat-containing protein (TAIR:AT2G22070.1); Has 46660 Blast hits to 13875 proteins in 226 species: Archae - 0; Bacteria - 4; Metazoa - 74; Fungi - 58; Plants - 45878; Viruses - 0; Other Eukaryotes - 646 (source: NCBI BLink).	YARNGRFKEALEMFCHMVSEGIGVDKYILTTIVSV CADYGLLSFGEQIHGLVHKIGHKFDIVFCSSLVDM YSKCGCLENAQIVFDKTTTRNTILWTAIINGNAFH GFGREAVKLFKLMIDDDIQPNEVTFVAVLTACSH AGLVDEGRNYFTVMXGIEHFTCMVDLFGRAGYL EEIKNFIYENNLSHLSSVWRAFLSSCRLHKNYDL GKWVSEKLLQLEPL
496	sp_207541z12186	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
497	sp_208330_izgx	0.923119	#N/A	#N/A	Polynucleotidyl transferase, ribonuclease H-like superfamily protein; FUNCTIONS IN: nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Polynucleotidyl transferase, ribonuclease H fold (InterPro:IPR012337); BEST Arabidopsis thaliana protein match is: thioredoxin family protein (TAIR:AT1G52990.1); Has 414 Blast hits to 414 proteins in 36 species: Archae - 14; Bacteria - 19; Metazoa - 0; Fungi - 5; Plants - 373; Viruses - 0; Other Eukaryotes - 3 (source: NCBI BLink).	DGRVVCPTLQAASFTKLVADYIEYSQRVFKPRPN FTRTSPARLSSPSPGYIKVNTDANIVSEAGVGLG ALLRDDSGAPVAAKVKRLKARWSVEQAEVEAVK YGMQLANRLGFTKVLMECDAVTVVQAINNNYKG YSPLLLYYEDIARMKLSFDTFQCSHXQVQQGCLP PHQVTLR
498	sp_208441z13982	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
499	sp_208921z15390	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
500	sp_209531z16277	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
501	sp_210441z17064	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
502	sp_211480_shfn	0.923119	#N/A	#N/A	uncharacterized protein	MSRAAPRTPALRTEPSAEHISIMPSTAAPKLGVE FRNFPKLLAEAQPSKSKSKSVPSPMSRAAPRTP ALRTEPSAEHISIMPSTAAPKLGVEFRNFPKLLAE AQPSKSKSKSVS
503	sp_211851z17738	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
504	sp_213401z18122	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
505	sp_213691z18153	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
506	sp_213691z18154	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
507	sp_213941z18231	0.923119	#N/A	#N/A	uncharacterized protein	#N/A
508	sp_213980_acik	0.931835	#N/A	#N/A	uncharacterized protein	MTHKDATTINLVQWYANDNFAQRKSKSICNSTLS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
509	sp_214250_arhi	0.9224193	#N/A	#N/A	O-Glycosyl hydrolases family 17 protein; FUNCTIONS IN: cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: carbohydrate metabolic process; LOCATED IN: anchored to membrane; EXPRESSED IN: 11 plant structures; EXPRESSED DURING: L mature pollen stage, LP.06 six leaves visible, M germinated pollen stage, LP.04 four leaves visible, 4 anthesis; CONTAINS InterPro DOMAIN/s: X8 (InterPro:IPR012946), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, family 17 (InterPro:IPR000490), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: O-Glycosyl hydrolases family 17 protein (TAIR:AT5G18220.1); Has 2719 Blast hits to 2656 proteins in 173 species: Archae - 0; Bacteria - 8; Metazoa - 5; Fungi - 96; Plants - 2598; Viruses - 0; Other Eukaryotes - 12 (source: NCBI BLink).	TYADYTPLGYGSSCHGLDANENISYAFNVYYQTQ YQKLEACNFRGLAKLTTQNISTGNCNLTIQIASSG LRTAAPAAFLLNVVLIYLFF
510	sp_214270_dtiz	0.9223119	#N/A	#N/A	RNA-dependent RNA polymerase family protein; FUNCTIONS IN: RNA-directed RNA polymerase activity; INVOLVED IN: posttranscriptional gene silencing; LOCATED IN: cellular_component unknown; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: RNA-dependent RNA polymerase, eukaryotic-type (InterPro:IPR007855); BEST Arabidopsis thaliana protein match is: RNA-dependent RNA polymerase family protein (TAIR:AT2G19910.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	AVNWDFSDEFIAARMVASGIPLDEPFLKHRLSEIA KEEMKSLKGGKLPVSGSFYLMGTADPTGTLNSG EVCVILYVFFFVISPAICCYMSRRCYH
511	sp_215551z18525	0.9223119	#N/A	#N/A	uncharacterized protein	#N/A
512	sp_215871z18557	0.9223119	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
513	sp_216290_qcaa	0.828333	#N/A	#N/A	FAD-binding Berberine family protein; FUNCTIONS IN: oxidoreductase activity, FAD binding, catalytic activity; LOCATED IN: endomembrane system; EXPRESSED IN: embryo, synergid; EXPRESSED DURING: C globular stage; CONTAINS InterPro DOMAIN/s: FAD-binding, type 2 (InterPro:IPR016166), Berberine/berberine-like (InterPro:IPR012951), FAD linked oxidase, N-terminal (InterPro:IPR006094); BEST Arabidopsis thaliana protein match is: FAD-binding Berberine family protein (TAIR:AT1G26420.1); Has 4385 Blast hits to 4167 proteins in 821 species: Archae - 36; Bacteria - 1918; Metazoa - 13; Fungi - 1639; Plants - 666; Viruses - 0; Other Eukaryotes - 113 (source: NCBI BLink).	ATFLSIIWIGILVSPILSVVVTNNNNIHLDAFIQCLE NESHVSDGYPFSEFILTPNNSSFDQVLHNRTNNL RYEQPTTRKPLLIVAPKSELNVPAVVTCARNNGL KMKIRSGGHDFEGISFTASFPFILLDMFNLNKVDI DIDTETAWVQSGATLGQIYYRIAEKSPIHAFPAGV CPSVGAGGHFSGGGYGVLIRKYGLSVDNIVDAR VVDVNGRILDRESMGEDFFWAILGGGGSSFGVIL SWKLNLVRVPKKVTVFNVNKDVNKPGVNHMLYK WQNIAHKLPNDLLIRVESSIYRTPENPYENPTIYA SFISMFLGDRENLVKIMNKSFPELGLTRKDCKEM SWIDSTMFWHRIPDGTPREILLNYDIKHSAYRKIK GDFVKKPIPEAALPQIWEKMLQTGQVYMEWTPY GGRMDEIDESAVPYPHRKGNLFMIFYDLLWIEDG VDEFYLNKARELHQFMGQFVSDSPRESVLNYVD LDIGSSWEKYFKANFPKLVMVKTRVDPTDFFSHE QSIPTLYKLLENKDFKRGYSNLLRLPMS
514	sp_000180_aycw	-0.819482	-0.85531	-0.855857	flavodoxin family protein / radical SAM domain-containing protein; FUNCTIONS IN: iron-sulfur cluster binding, oxidoreductase activity, FMN binding, catalytic activity; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Flavodoxin/nitric oxide synthase (InterPro:IPR008254), Radical SAM (InterPro:IPR007197), Wyosine base formation (InterPro:IPR013917); BEST Arabidopsis thaliana protein match is: P450 reductase 2 (TAIR:AT4G30210.2); Has 3108 Blast hits to 3089 proteins in 883 species: Archae - 191; Bacteria - 976; Metazoa - 668; Fungi - 593; Plants - 205; Viruses - 0; Other Eukaryotes - 475 (source: NCBI BLink).	METTPSLACANKCVFCWRHHTNPVGKSWQWK MDDPVEIVNTALDLHTKMIKQMKGVPGVLEERLS EGLSPRHCALSLVGEPIMYPEINTLVDELHRRRIS TFLVTNAQFPEKIKLLKPITQLYVSVDAATRDSLK AVDRPLFGDFWERFLESLTSLNEKHQRTVYRLTL VKGWNAEDVDGYFKLVSLGQPDFIEIKGVTYCGS TATSKLTMENVPWHADVKAFSEAVALRSEGEYE VACEHEHSCCVLLAKVDKFKINGQWFTWIDYEKF HDLVASGEPFKSSDYMAPTPFWAVYGAVEGGF DPQQARFRKERNHKSKLQSK
515	sp_000360_ejow	#N/A	-0.8511571	#N/A	E3 Ubiquitin ligase family protein; FUNCTIONS IN: zinc ion binding; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841), E3 Ubiquitin ligase (InterPro:IPR022170); BEST Arabidopsis thaliana protein match is: E3 Ubiquitin ligase family protein (TAIR:AT1G59560.1); Has 1960 Blast hits to 1959 proteins in 225 species: Archae - 0; Bacteria - 0; Metazoa - 1182; Fungi - 34; Plants - 350; Viruses - 140; Other Eukaryotes - 254 (source: NCBI BLink).	MSAGDQAAVAIAAQLLLAGDGACLGLALGYVAV RSLTKFAATSSALRKIGDAPSVQVSDLRFLIKEDS GGGDEGEFRNSEVLVVVRGLVASKSAVKGSWKS LWPDVLVSHESGDRGVVLQRSQKCIYNEWRGF FGWVYPDLRAVLSKAWREQGTTSTRTVPFVLVDG RQQTQSEYVVVDLDGSNHPLPLTTIYQIHMQPIEA SPYTFLQALFGHEYPVGVLDEEKILPLGKEVTAV GICSSKSGALEIKSCRDLPCFLSELSKDQILVQLS FRTKILLWSGIFIGSVSIGILGYAIVRNWNKWKQW RVQRRAEQQNHPAAAESDNSSPESEVEDVDED VGDVPDGQLCVICLMRRRRSAFVPCGHLVCCRR CAFSVERDLAPKCPVCRQSIRGSVRIYDA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
516	sp_000640_ugxj	-0.884324	#N/A	-0.922388	Tetratricopeptide repeat (TPR)-like superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT5G39710.1); Has 28218 Blast hits to 10356 proteins in 245 species: Archae - 1; Bacteria - 15; Metazoa - 229; Fungi - 251; Plants - 27094; Viruses - 0; Other Eukaryotes - 628 (source: NCBI BLink).	MSSTIFRRALFPRATTTTFSFSSISSSAHASPSPP STSSVATAVSILTHHRSKSRWSALLSLFPHGFTP PQTSQILLGIRNNPHLSLRFFLFTTRHSLSSLHSY ATIIHILARGRHKTRAKTLIKSAIREFDLLGADDSQ VFSGKKCKVFRALAETYRVCDSAPFVFDLLILSLL EMNKLDPCIDIVRMLHSRAIYINASTCIDLIVSASKI RGCFFGYDVYREVFGYDDSGFVDHDKKVNCKLV PNVHIFNELMLSFHRNGLLDMVEEVWIEMVKSNC EPNSYSYSILMAAYCEDNDMEKAEKVWEELRGK GLNPDIVGYNTLIGGYCRIGEVRKGEELYKEMGL RGLNGSFVTYEHLIGGYCLCGDYDSAMLLYNDM RRKGFNPNSLIVDSLVRILCGKRSVSEGLEFFRVA MKNADFVVSESSYQLLIEGLCDEGMTKEALTLQA QMAGKGFKPNSQIYDAFIDGYMKIGNTDEAENLR KEMMDK
517	sp_000790_mejd	#N/A	-0.864825	#N/A	beta-amylase 3 (BMY3); FUNCTIONS IN: cation binding, beta-amylase activity, catalytic activity; INVOLVED IN: cellulose biosynthetic process, carbohydrate metabolic process, polysaccharide catabolic process; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase, family 14 (InterPro:IPR001554), Glycoside hydrolase, catalytic core (InterPro:IPR017853), Glycoside hydrolase, subgroup, catalytic core (InterPro:IPR013781); BEST Arabidopsis thaliana protein match is: chloroplast beta-amylase (TAIR:AT4G17090.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MLIPKNRVSFENSNHRVGFTLKAVQGDAIVAEEK ANSILLTSCQSVGGVRLFVGLPLDTVSQYNTVNH ERAMTAGLKALKLLGVDGVELPVWWGIVEKDNM GIYDWSGYLAIAEMVQNLGLKLRISFCFHASQEP KISLPDWVSRVGEADPNIYFMDRSGRQYKDCLS LAVDDLPVFDSTSPIQVYQNFLESFNAIFSPFMGS TITGITVGLGPDGELRYPSFHDATTTALNNQLGV GEFQCYDKYMLSHLKQYAEETGKPMWGLGGPH DVPKYDESPNANNFFKDDGGSWETPYGDFFLS WYSHQLVEHGNRILSLASTVFGEIPTTISGKLPLV HSWYKTRSHPCELTAGFYNTVNRDGYESVVEMF AKNSCRIILPGLDLSDAHQQNGSLLVQIKDSCTKQ GVGMYGENSSVCRVNNGFEKIKSNLSEEKAVVS SFTYQRMGADFFSPKHFPIFTAFVRAVNQLDLDL DVDDMPDTTTTRAVETIHMPSEALKLQTA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
518	sp_000880_wxck	#N/A	-0.873388	#N/A	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: DNA helicase activity, nucleoside-triphosphatase activity, nucleotide binding, ATP binding; LOCATED IN: nucleolus, chloroplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: TIP49, C-terminal (InterPro:IPR010339), ATPase, AAA+ type, core (InterPro:IPR003593); BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT3G49830.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAELKLSESKDLTRIERIGAHSHIRGLGLDSTLEP RLTSEGLVGQLPARKAAGVILQMKDGKIAGRGV LIAGQPGTGKTAIAMGLAKSLGHETPFAMLSGSEI FSLEMSKTEALMQAFRRAIGVRIKEEAEVIEGEVV EVQVDRPAVAGAASKTGKLTLKTTDMETVYDLG TKMIEALSKEKVQSGDVIAIDKASGKITKLGRSFS RSRDYDAVGPHTKFVQCPDGELQKRKEVVHCVT LHEIDVINSRTQGFLALFTGDTGEIRAEVREQIDT KVAEWREEGKAEIVPGVLFIDEVHMLDIECFSFLN RALENEMSPILVVATNRGHTRIRGTNYKSPHGIPI DLLDRLLIIATQPYSEEDIRKILEIRCQEEDVEMTE DAKVLLTKIAVETSLRYSINLINSSALACLKRKAKS VEMEDVSRVYQLFLDVKRSTQYLMEYQNQYMFN DVSGGDLNEDDNTAMLS
519	sp_000960_ucxn	-0.886376	#N/A	#N/A	Ribosomal RNA adenine dimethylase family protein; FUNCTIONS IN: rRNA methyltransferase activity, rRNA (adenine-N6,N6-)-dimethyltransferase activity, rRNA (adenine) methyltransferase activity; INVOLVED IN: rRNA modification, rRNA processing; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal RNA adenine methylase transferase, N-terminal (InterPro:IPR020598), Ribosomal RNA adenine dimethylase (InterPro:IPR011530), Ribosomal RNA adenine methylase transferase, conserved site (InterPro:IPR020596), Ribosomal RNA adenine methylase transferase (InterPro:IPR001737); BEST Arabidopsis thaliana protein match is: Ribosomal RNA adenine dimethylase family protein (TAIR:AT5G66360.2); Has 9871 Blast hits to 9860 proteins in 3038 species: Archae - 214; Bacteria - 6519; Metazoa - 271; Fungi - 146; Plants - 120; Viruses - 0; Other Eukaryotes - 2601 (source: NCBI BLink).	MAGGKIRKEKQKGSNHFQGGIPFHKSKGQHILKN PMLVESIVQKAGIKPTDIILEIGPGTGNLTKKLLEA GKKVIAVELDPRMVLELKRRFQGTPFSTRLEVIQ GDVLKCELPFFDICVANIPYQISSPLVFKLLDNQK QNFRCAIIMFQREFAMRLVAQPGDKLYCRLAVNT QLLARVFHLLKVGKNNFRPPPKVDSSVVRIEPRR PLPIPVGKLKEWNGLLSLCFSRKNKTLGSIFGQK KVLEMLEKNYRTLESLGQLKDNDSEIIDMSILNED VSMDVVENGKADDEEDDQMEVEEEGEDDADSD NNEMQINDSGFKGKVIGVLKQYDYADKRSSKLTQ NEFIHLLSVFNQAGIHFTSTK
520	sp_002390_ttfp	-0.881844	-0.852507	-0.883079	uncharacterized protein	MSSLPSNPSSTVEKEKEKEKEKEKEKEKEKE KEEKEKSKSKSKSSNRLTKEEQFQAKNQTFPPL ERMEINRGFINTTSLVEGPPRSISLVIGFSINIDIWF RGDWDGKDLNDIALCQWQLSWDAMDPEVARIT RGPFWKNFSDPTDILSDDEAHYKHLLAGYDQPL SAHFEGPDGSCITMYYAILLHAVQAWVQVEVIDV NVVPTAVTYTVHGNIFASYGHFESDNDSDIKDYC KTTLLKTGDLESVKVERWSKIKLSRSIVDTLGGM MTLLFG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
521	sp_003590_tmqe	#N/A	#N/A	-0.884475	Protein of unknown function (DUF1664); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1664 (InterPro:IPR012458); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF1664) (TAIR:AT2G02730.2); Has 199 Blast hits to 190 proteins in 29 species: Archae - 0; Bacteria - 14; Metazoa - 2; Fungi - 2; Plants - 161; Viruses - 0; Other Eukaryotes - 20 (source: NCBI BLink).	MAMQAGLGVSRLLLLVGAGYTGTVLVKNGKLSDT IGELQGLVKGLENSDDSANADSEIAKQVRKLAQE VRNIASSRQITVVNGGSGGTDFSSLILPAASLGAL AYGYMWWKGLSFSDLMYVTKRNMATAVANLTA NLDNITDRLSKVKLHLTQRIENLDGKVDEQREMS KQIKDEVCKTRSEIDNVNSQLYALKDTVNSLGKD MESFEAKQMLTLSGVKYLCDYVDGRTSKAPKIS QHLQEQVKHGKIRAQLSSSNALGLQGLKDIAESV NVTRSIKEIE
522	sp_003620_widu	#N/A	-0.906736	-0.959723	Cystathionine beta-synthase (CBS) protein; CONTAINS InterPro DOMAIN/s: Cystathionine beta-synthase, core (InterPro:IPR000644); BEST Arabidopsis thaliana protein match is: SNF1-related protein kinase regulatory subunit gamma 1 (TAIR:AT3G48530.1); Has 879 Blast hits to 799 proteins in 237 species: Archae - 117; Bacteria - 170; Metazoa - 187; Fungi - 63; Plants - 180; Viruses - 0; Other Eukaryotes - 162 (source: NCBI BLink).	MDTATTVGYDAYFETVQIQSRKKLPISLQETLTSA FASIPVSSFPKVPGGKVIEISAETSVSDTVKVLSE CNILSAPVTDMEKANSNDWKERYIGIIDYSAIVLW VLQGAELAAVALSATSATAAGVGAGAIGTLGAVA LGATGPVAVAGLAAAAVGAAVAGGLAVDKGMG KDAPTAASELGDDFYKVILQEEPFKSTKVREVVK SFRSTPFLPVSTESNMLTVLLLLSKYRMRNVPVIE PGKAFMKNFITQSAVVRGLEGCRGLDWFDCIAA HPISDLGLPFMSRDQVVSIDSSELILEAFKRMKDH HIGGLPVVEGPKRKIVGNLSIRDIRYLLLQPNLFSN YRHLTVNDFMKTVTSMCQEAGRIMPPITCKSDTT LGDVIHSLASKSVHRIYVVAEEDEVIGVITLRDVIS CFIYEPPNYYDEYLGYSAEVLLDR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
523	sp_003800_wscm	-0.3937335	#N/A	#N/A	SAC3/GANP/Nin1/mts3/eIF-3 p25 family; CONTAINS InterPro DOMAIN/s: SAC3/GANP/Nin1/mts3/eIF-3 p25 (InterPro:IPR005062); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MTSEGGSATQTVAAATADPNSIETRAEVNQGRAH QYYPSAAGATPVAWPIHGGDNSSLGNGTVQNV SYQYNQHQPATVQDATNVPSITSTSTASGTDNA SHMYTSYSTYSSSTNPYGYASSEYTNYYSGYQQ QSTEPSGYQQQSTQPSGYQPQSTQPSGYQQQS TQPSGYQQQSTQPSGYQQQSTQPSGYQQQST QSSGYQQQSTQPSGYQQQSTQPSGYQQQSTQ PSGYQQQSTQPSGYQQQQSSQPYTQATGTYQNS GSPYQPISSFQNTGSYTGSTTYPSTYYNFGDYQ TTGAYNGSSYSSQTNLWSQGNYANYPHQYESY SNDTSSAYTSSNVAATTDYQQQYQQWADYYNS QTEVTCAPGTENTSASATAALSCPIPGVSSGYAN PVSQPQATYTLAWKPETDPSALSSHQTSSTAGT YNDYWKHAAQAYQNQHASNLQSNFQKPLDVSH SHPSVQDQQQTLATQPQLPNSLYTEVQQTYPQP LQPHTPADTLRVSNLQIPTNPRIASNLALAMPKTI KDTVTSGVTAKPAYVSVSLPKPTNKVLSSSASDS NPKPGIFPAALRGYVERALARCKDEKQRAASQE VMKEIITKATADGTLYSRDWDTEPLPPLPDTSTLN IEQSAAPVLSLSAFKRSPSRRAKSRWEPLPEDKP IEKLTSTYQDSARPVWANYNGRDHKSVAGKSER KDATFSRFSTVDQTSTSRSAHRPNKRQRFSNDL SATENGDASSDSDKEQSLTAHYSGAITLANSPEE KKRRESRSRRFEKGNGQRAENRFRPKNAGVGN YLYSRRATALVLSKTFEDGGSRAVEDIDWDALTI KGTCQEIEKRYLRLTSAPDPSTVRPEEVLEKALH MVQSSQKNYLYKCDQLKSIRQDLTVQRIFNELTV KVYETHGRLALEFGDLPEFNQSQSQLKSLYGEGI AGCHMEFSAYNLLSISLLSNNRKDLLSLMSRLSD ETKKHEAVKHALSVRSAVTSGNYVLFFRLYKSAP NLNTCLMDLCVEKMRFEAVKCMSRTYRPTIPVSY VAQVLGFASAPEGEGGDEGDVDGLEECMEWLK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
524	sp_004140_rkiz	#N/A	-0.955318	-0.848322	CAB UNDEREXPRESSED 1 (CUE1); FUNCTIONS IN: antiporter activity; INVOLVED IN: transport; LOCATED IN: mitochondrion, chloroplast, plastid, plastid inner membrane, chloroplast envelope; EXPRESSED IN: callus; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF6, transmembrane (InterPro:IPR000620); Tpt phosphate/phosphoenolpyruvate translocator (InterPro:IPR004696), Protein of unknown function DUF250 (InterPro:IPR004853); BEST Arabidopsis thaliana protein match is: phosphoenolpyruvate (pep)/phosphate translocator 2 (TAIR:AT3G01550.1); Has 2791 Blast hits to 2789 proteins in 409 species: Archae - 9; Bacteria - 295; Metazoa - 571; Fungi - 421; Plants - 1194; Viruses - 0; Other Eukaryotes - 301 (source: NCBI BLink).	MQSSAISISPSLPHLKPLQNPSFQHRRLPPLNSFS VRPSLSSAITSVPRLTIVRSSSEEDSSFNPSRRSS RFDGILFSSPSLNFSGKRSSQDDFAVKATSSSLPE SADSDEAPQTMSQTLEIGILFVLWIVINIYFNIYNK QVEFDF
525	sp_004220_jsph	-0.936583	-0.912767	-0.837574	unknown protein; Has 1524 Blast hits to 1298 proteins in 225 species: Archae - 9; Bacteria - 84; Metazoa - 474; Fungi - 184; Plants - 98; Viruses - 17; Other Eukaryotes - 658 (source: NCBI BLink).	MGKKKFIDKKNSATFQLLSRDSSDPVYIEEPEND RVFVFRVDNNEVSVNGIDEEHDVESADDVGGSEV FRGPTPLTDRVRREILELGFPDDGYNYLTHLREIR NTGGGSAYFQNPKAELHQLPHDVKAYDASRVRI KEASEDTSPKTLYNVAEKTVAVRVGKAVDPEIAA LLEDNESDFGSDIDDLEEDFVVGANHSEETGDVY EDTTLSLVNNPQEEVLNQSDTLISCDIQTEVASNS TVEVVQSVNEKSRTQRLLDQQFDMLESQEYGSD SDDDYDNYVAEQEISLAERLKNVLKDREVDDLEF NENYKAPAEILRRCAEYAEKYENEDGNDKVVVIV EESSDESEKWDCETFVSTYSNLDNHPGRIGVPG RIRQKILTETISGVLKSGSDKIISLGGKARLPVDYL PHGKKESKETTKVVAKTEQPKRKPHGQETKEEK KERKAAVKEERREAERRLKKDLKELYRDEAKLAQ KVAAITGPSSIHLG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
526	sp_004890_dmip	#N/A	-0.876342	#N/A	glucose-6-phosphate dehydrogenase 2 (G6PD2); CONTAINS InterPro DOMAIN/s: Glucose-6-phosphate dehydrogenase, C-terminal (InterPro:IPR022675), Glucose-6-phosphate dehydrogenase, active site (InterPro:IPR019796), NAD(P)-binding domain (InterPro:IPR016040), Glucose-6-phosphate dehydrogenase (InterPro:IPR001282), Glucose-6-phosphate dehydrogenase, NAD-binding (InterPro:IPR022674); BEST Arabidopsis thaliana protein match is: glucose-6-phosphate dehydrogenase 3 (TAIR:AT1G24280.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MATFSPSHCASSSSSFYSSYSSSKSRNGRLGMV KISSISKQSAPKISVSKQQQSGRNPKSNAVFMQD GAVATKIIPVEGENVSSVETLKDGILSATSSKESN DLSGFDVRKDESTVSITVVGASGDLAKKKIFPALF ALYYEDCLPKHFTIFGYARSKMTDDELRTMVSKT LTCRIDKRENCGEKMEEFLKRCFYQSGQYDSRE NFVELDKKLRQREAGRIANRLFYLSIPPNIFIDAVR CASLSASSANGWTRVIVEKPFGRDSESSAALTN GLKQYLKEDQIFRIDHYLGKELVENLSVLRFSNLIF EPLWSRQYIRNVQLIFSEDFGTEGRGGYFDHYGI IRDIMQNHLLQILALFAMETPVSLDAEDIRNEKVK VLRSMRPLQLDDVVIGQYKSQTRGGVTYPGYTE DKTVPKDSLTPTFAAAALFIDNARWDGVPFLMKA GKALHNRSAEIRVQFRHVPGNLYARNFGSNLDQ SMNELVIRVQPDEAIYLKINNKVPGLGMRLDRSTL NLHYAARYAKEIPDAYERLLLDAVEGERRLFIRSD ELDAAWALFTPLLKELEQKKVNPEYYPYGSRGP VGAHYLAARYNVRWGDLGSEP
527	sp_005030_fqaq	#N/A	#N/A	-0.8371138	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 12 growth stages; Has 16 Blast hits to 16 proteins in 8 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 16; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MEFQNLIKDKKFWFASVLIAWAAGLQGHMMWL QRQDAFKQKFGDLNESDDQKNLN
528	sp_005450_ptpz	-0.8833332	-0.886698	#N/A	unknown protein; Has 20 Blast hits to 20 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 20; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MEEDWKEGVKRGGCERLQNALYDCHRRMPAG PARNSGCRHLNRAFTDCLVAVICPSEIEAVRSLC SSGGTAMKRSQCQEARLSLSVCLSQHQDQDHP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
529	sp_005830_iiwq	-0.8844412	#N/A	#N/A	pescadillo-related; FUNCTIONS IN: transcription coactivator activity; INVOLVED IN: cell proliferation; LOCATED IN: nucleolus, intracellular; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pescadillo, N-terminal (InterPro:IPR010613), BRCT (InterPro:IPR001357); Has 503 Blast hits to 494 proteins in 230 species: Archae - 0; Bacteria - 0; Metazoa - 173; Fungi - 168; Plants - 48; Viruses - 0; Other Eukaryotes - 114 (source: NCBI BLink).	MPKHYRPAGKKKEGNAAKYVTRSQAVKGLQVTL PLFRKMCIWKGIFPRDPKKKIKGNHQTYYHTKDIS FLHHDPLVDIHRDIRAQDKKIKKAYSKKNKPRAVR LEEHKPTYNLDRLIIERYPRFIDALGDLDDCLTMV HLFAALPTLQNEKVEVKHHSCRRLSHEWQAYIS RTHKLRKTFISVKGIYYQADVDGQKITWIAPHSIQ QVLTEDVDFNVMLTFLEYYETLLAFVNFKLYHSIN VKYPPMLDPRLEALAADLYALSRYIASRSKEQDN HELASQVDESQLTLAQLQHQLAPSEPGALMHLIE EAETVDEADEETKECRSLFKDTKFFLGREVPREP LLFVITAFGGTISWEGDGAPFEESDPSITHQIVDR PTQNHMFLSRAYIQPQWVFDSVNARIILPTEDYV VGRVPPPHLSPFVDNEAEGYIPDYAQKIKQLQDA ARIMPVPEVGNEHQDNTQNIVLDGVADRIEANKA AEMKKKMSAGQKQFHDELKMEVEGVSPATGTIK NVKVTTDPDESDKEQISEHPDMSEMTMSRRKRG LYESIRINKERKKAKVDVIESRKTRNIKESQK
530	sp_006460_eyhq	-0.8755336	#N/A	#N/A	ARM repeat superfamily protein; FUNCTIONS IN: protein transporter activity, binding; INVOLVED IN: intracellular protein transport, protein import into nucleus, docking; LOCATED IN: nucleus, chloroplast, nuclear pore, cytoplasm; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Importin-beta, N-terminal (InterPro:IPR001494), Armadillo-like helical (InterPro:IPR011989), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: ARM repeat superfamily protein (TAIR:AT3G08947.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MVMEITQLLLSAQSVDGQVRKHAEDSLKQFQDQ NLPNFYLSLAGELASEEKPADSRKLAGLILKNELD AKEQHRKFELVQKWVSLDAGLKNQIKTFLLQSLT SPVHDARSTASQVVAKVAGIELPQKEWPELIGSL LGNIHQLPPHVKQATLETLGYLCEEVSPDVVDQD HVNKILTAVVQGMNANETNNEVRLAATHALYNAL GFAHANFNNDMEFDYIMRVVCEATMSPDLKIRQ AAFECLVAISSTYYEKLAPYIQDIFSITAKAVREDE EPVALQAVEFWSSICDEEIDILEEYGGDFTGDSEI PCFYFIKQALPALVPMLLETLLKQEDEQDQDEGA WNIAMAGGTCLGLVARTVGDDIVPLVMPFIEENIT KPDWRQREAATYAFGSILEGPSPDKLTAIVNVAL NFMLTALTKDPNSHVKDTTAWTLGRIFEFLHGSA METPIITPANCQQIVTVLLQAMKDTPNVAEKACG ALYFLAQGYEDVASSSPLTPYFQEIVQCLLTVTH REDAGEARLRTAAYETLNEVVRCSTDETAPLVLQ LVPVIMMELHKTLESHNLSPDGKQGELQGLLCG CLQVIIQKLGSSEPTRAVLLQYADQIMQLFLGVFA CRSATVHEEAMLGIGALAYVTGPDFAKYMNEFYK YLEMGLQNFGEYQVCAVTVGVVGDLCRALEDKIL PYCDGIMTQLLKDLSSNQLHRSVKPPIFSCFGDIA LAIGENFEKYLMYGMPMLQSAADLSAHTAGADD EILEYTNLLRNGILEAYSGILQGFKNTSKTQLLLPH ARHILSFIDSIYQEKDMDDAVMKTAIGYLGDLADT LASSAGSLIQQFPTCRELLDEGLSSDDHSIKQSAE WAKLAVSRVISI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
531	sp_006590_fpjx	#N/A	-0.89987	#N/A	uncharacterized protein	EYLTFSIKFLKWGDPTNFKYREELKFVHMSTFFS ASRAATLGRAAARRRGERARDACTKVAGRAAPKE RE
532	sp_006900_anhc	#N/A	-0.907164	-0.90036	MINICHROMOSOME MAINTENANCE 3 (MCM3); FUNCTIONS IN: nucleoside-triphosphatase activity, DNA-dependent ATPase activity, DNA binding, nucleotide binding, ATP binding; INVOLVED IN: cell proliferation, DNA-dependent DNA replication initiation, DNA unwinding involved in replication; LOCATED IN: nuclear chromatin; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), Nucleic acid-binding, OB-fold (InterPro:IPR012340), ATPase, AAA+ type, core (InterPro:IPR003593), DNA-dependent ATPase MCM (InterPro:IPR001208), DNA-dependent ATPase MCM, conserved site (InterPro:IPR018525), MCM protein 3 (InterPro:IPR008046); BEST Arabidopsis thaliana protein match is: Minichromosome maintenance (MCM2/3/5) family protein (TAIR:AT4G02060.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MWIKKAVILLMLGGTEKNLKNGTHLRGDINMMMV GDPSVAKSQLLREIMNIAALAISTTSRGSSGVGLT AVVTSDQEMGKEG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
533	sp_007210_sudc	0.8887227	#N/A	#N/A	Heat shock protein DnaJ with tetratricopeptide repeat; FUNCTIONS IN: binding, heat shock protein binding; INVOLVED IN: protein folding; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide TPR-1 (InterPro:IPR001440), Tetratricopeptide-like helical (InterPro:IPR011990), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734), Heat shock protein DnaJ, conserved site (InterPro:IPR018253); BEST Arabidopsis thaliana protein match is: Heat shock protein DnaJ with tetratricopeptide repeat (TAIR:AT5G12430.1); Has 24543 Blast hits to 21405 proteins in 3009 species: Archae - 131; Bacteria - 8314; Metazoa - 5854; Fungi - 2610; Plants - 2808; Viruses - 9; Other Eukaryotes - 4817 (source: NCBI BLink).	MPPPAVSVDSRAPSDSNFNPQNGSTTHFNDCSS FNTGSSSTSFPAVGRGSSSTRAKPRLVKLRKSS KSSSLSSEADLDDSGINPFRSVSSLNTDASNCTN SSSDSFVSRFRNVGFVFGVSNSIDNNNINSNNKV NDGGKVAAEASVFGFRDSNLHKQEFSFSENGTE RDNNCTENVAFRFGATAKDQIGVSNVKCGENGV EFELLDELQKLKIDKERCTAEDKVDELNNVDTLK GGDRFFVFKTGGVSYDECGGSSVPLRSDVDATN LSGNGKGNNESKMGAFGYRYVDDSAFVFGSSG NFSKMGNSSSQSSSVPAENDGARIKEDFLSPSF SFSSLDADAGNAKSEACSSSSERKVEFSSSGLP EGSGVSFSEFELPQFDAPLSFSANFTSGLGRKPE SRGLKDKGLRKTKGKSKQSKPTKLQPVQTPLSS EGVTQGQEASECYSPMDFSPYEDTTNNDVPLTG DHETYQTKVTTGNSLDVHTDSELNDSSTDIKSDN HRGNEAQFSFSSDKASRDEENITFTSPTSAQNH VSSPKRLLKKKYRMKPASISSNASRVQGSSLDSS PLLSKLQGAGGYFFASASKDQHETEAGVEHDCE EFSAAEKACEKWRRRGNEAYEKGELSKAEDFYT WGVNCIPPSDASGAAFRPLVRCYSNRAATRMAC GRVREAIQDCLTAVKLDPTFHRAQIRAGNCYLLL GEVEEAMQCFGKYIGDGNVLCLDRKAVISASDGI QKSQKVLQSTSRSAELLKLKTSEAATDALQHIAE ALAISINSEKLLKMKAQALCTLSRHDEAIALCQKTL DFAKKNFTAAIFGNYEVDGSSNDRLWRYRILSKC QFHLGNLEVALDLLEKQENKTEGISDPLSATIREL LKQKSAGNSAFHSGNHAEAVEHYTSAIMSSMDA RPFAAICFCNRAAANQALGQITDAISDCSLAIALD ENYAKAVSRRATLHEMIRDYKHAAADLQRLSSFL QKPSVEKSKQSGSGNGKELRQVKQRLSMMEEE AKRDTPLDLYLILGVKKSDTVSEIKKAYRKAALRH HPDKAAQSRVKSGDVGEEGRHWKDIAELVHKDS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
534	sp_007320_xjcx	#N/A	#N/A	0.8222227	eukaryotic translation initiation factor 3C (EIF3C); FUNCTIONS IN: translation initiation factor activity; INVOLVED IN: translational initiation; LOCATED IN: cytosol, nucleus; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Winged helix-turn-helix transcription repressor DNA-binding (InterPro:IPR011991), Proteasome component (PCI) domain (InterPro:IPR000717), Eukaryotic translation initiation factor 3 subunit 8, N-terminal (InterPro:IPR008905); BEST Arabidopsis thaliana protein match is: eukaryotic translation initiation factor 3 subunit C2 (TAIR:AT3G22860.1).	MSSRFWDQGNSDTEEEESDYDDVDEPVADEAG AKAKYGPGDRYKEDNGNDSDDSDDDHKRVIIPGK DKRFQEMTATVEQMKNAIKINDWVSLQESYDKIN KQLEKVMRVAETGKVPNVYIKALVLLEDFLNETLS NKEAKKKMSTSNAKALNKMKQNLKKNNKQYED MITKFRENPESEDEKDEGDDSSGTEFTESEGED EDTDKEDDKGEGDDVEDDVENIGGWQTPGKNN RLIDIKLLKNPSEITWDTVDKKVKEIVAGRGKKGT GKVELVERLTFLTRIAKTPAQKLEILFSVVSAQFDI NPSLSGHMPINVWKKCVQNMFVILDILTQYPNILV DDSVEPEENETQKGPDHKGTIFVWGNLVAFLERI DVEFFKSLQCIDPYTREFVERLRDEPLFLVLAQNV QEYLEKSGNLKAASKIALKRVEFVYYKPQEVYDA MRKLAEQSEEGNDEAAGEEFKDAGLGRGPPLFV VTPELVPRKSTFPERCRALMDSLVSLIYQFGDER TKARAMLCDIYHHAIMDEFPIARDLLLMSHLQDTI QHMDISTQILFNRAMAQLGLCAFRIGLTSEGHSC LSELYSAGRVRELLAQGFSQSRYHEKTPEQEMK ERRQQMPYHIHINLELLEAVHLTCAMLLEVPYMA ANTFDSQRRIISKTFRRLLEASERQTLRGPPENV RDHAMAATRGLTQGDFQKAYDVITSLDVWKLLK NKESVFEMLKSKIKEAALRTYLFTYAVSYSSLSVD QLSKMFELSDPEIHSIVSKMMVNEELRASWDQP TQCIIFHEVEHTRLQALAFQLTEKLSFFAENNERA MEAKLGGGSLDGLPLRRRDGPDYAQVTSTGGG GGGGRWHDNSYNQGRSGGRSGYGGKSMNPG QFSGGGYGGHSRGGGGGGGSYGRARGQHNDS SARMVSLNRGAR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
535	sp_007710_askk	-0.922623	-0.661548	-0.899248	basic leucine-zipper 42 (bZIP42); FUNCTIONS IN: DNA binding, protein homodimerization activity, protein heterodimerization activity, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; EXPRESSED IN: inflorescence meristem, flower; EXPRESSED DURING: petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Basic-leucine zipper (bZIP) transcription factor (InterPro:IPR004827), bZIP transcription factor, bZIP-1 (InterPro:IPR011616); BEST Arabidopsis thaliana protein match is: basic leucine-zipper 43 (TAIR:AT5G38800.1); Has 1669 Blast hits to 1669 proteins in 135 species: Archae - 0; Bacteria - 2; Metazoa - 133; Fungi - 19; Plants - 1463; Viruses - 0; Other Eukaryotes - 52 (source: NCBI BLink).	MSSNSTSDESEEMHHQQGKMMIDDRRQRRMIS NRESARRSRMRKQRHLDELWSQVIRLRNENNNL IEKLNHVSESHETILQENVRLKQEASDLRQLINGM QLATTFNGLRDLEDIPCTTAHLRAESNSTHSITTS STDLLM
536	sp_008030_ekix	#N/A	-0.667436	#N/A		

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
537	sp_008300_wwdn	-0.864942	-0.859241	#N/A	Ring/U-Box superfamily protein; FUNCTIONS IN: zinc ion binding; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein (TAIR:AT5G04460.3); Has 608 Blast hits to 493 proteins in 135 species: Archae - 2; Bacteria - 207; Metazoa - 113; Fungi - 27; Plants - 119; Viruses - 0; Other Eukaryotes - 140 (source: NCBI BLink).	MAIAGLHNVPVIESSILRESQSSSARIGNPGSISTR ASSIRQMWLELEDECVVNRVRDRGRLRLQQERN DGLSNCQVGEGRGGLEDSSESENGSVINSESLT GPLNEQYDRQSLASEQSQDFGLVERARVRRIFQ EWKNTGVANQATNVPHRNRGVRSQWLGETER ERVRVVREWIQMTSQPREDQVSEVGSPIERVRD GQVVDHNESQVENNRRPIRRLCGRQALLDLLAK KEQERRQELQGLEVTQPVSRFAHRNRIQSLLRIR CFQNMRSAESRRPSSTAESELGLLRQRQTVSGL RGGFLSRLDNVQVQASDLSGSSLENSPHSFRDD QTEANSANEVVDGAHVPNNPANGFLDGADEQN QPANEVLSGAHEQNNPANENLDVICEQRNVEGD ANVNNPREYSGNTNFQELEVQVETRLEVASDEE NISETATSGPADGFSDEGYLLMNSQEPSRSGSE EFGEIEGTHELSRHFNTHREEINVLNQLDNEVHE SEEVNDQDFSIQVEDWPEQVTTYNESGWQGNV LSNQWRDGDQAETNRQHVDGEEHFDGALESEV GEQNDAQEDYWHESGSLETPREWLGMPSTSG VTSGRLDTYYFSDDDNVHHMELRELVNRRRVSS LLHSDFRESLDQLIQSYVERQTNVPEDWELHEM SLPPFTTQNQDQLSTGHSEESEPIHVEETQRTLP RPPPLPQAHWGHHANWSRHESRQRPGTEWEII NDLRIDMARLQQRLNNMQRMLEASMDMQLELQ RSVRQEVSAALNRSATPSELFENVLPKDSFKWD CVRKGICCICSNTNIDSLLYRCGHMCTCTKCADLL VQGDGKCPMCQAPVIEEGLYRNILKLSKLGQSIS LQMPVLTKSPVFLSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
538	sp_008750_cisf	-0.942767	-0.960379	-0.904767	Nucleolar GTP-binding protein; FUNCTIONS IN: GTP binding, nucleotide binding; LOCATED IN: membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: GTP1/OBG (InterPro:IPR006073), Nucleolar GTP-binding 1 (InterPro:IPR010674), NOG, C-terminal (InterPro:IPR012973); BEST Arabidopsis thaliana protein match is: Nucleolar GTP-binding protein (TAIR:AT1G10300.1); Has 9379 Blast hits to 9179 proteins in 2035 species: Archae - 370; Bacteria - 5195; Metazoa - 1143; Fungi - 469; Plants - 325; Viruses - 0; Other Eukaryotes - 1877 (source: NCBI BLink).	DEFPRLDDIHPFYGDLLHVLYNKDHYKLALGQVN TARNLISKISKDYVKLLKYGDSLYRCKCLKVAALG RMCTVLKRISPSLAYLEQVRQHMARLPSIDPNTR TILICGYPNVGKSSFMNKITRADVDVQPYAFTTKS LFVGHTDYKYLRYQVIDTPGILDRPFEDRNIIEMC SITALAHLRAAILFFLDISGSCGYTIAQQAALFHSIK SLFMNKPLIIVCNKTDLLAFDNIPEEDKKLVMDMK AEALKTVIGQGGEATNEEEGVLLTMSTLTEEGVM AVKNAACDRLLDQRVELKMKSKKINDCLNRFHVA MPKPRDQKERPACIPQGVLEAKAKEAAEKAEKE HRITERQLEDENGGAGVYSASLKKHYILANDEWK EDIMPEFLDGHNVYDFVDPDILHRLEELEREEGL RLADEAEEDFEMDGEELTVEEKEALAEIRRKKSIL IQEHRLKKSTAGSRAVVPRKFDKDRKFTTERMG RQLTKIGIDPSAAILRARSKSRGQKRQRSPEVGG EGDVMDMDVEQPNNKKLRALSRPRSRSLSRPP GEVVPGEGFKDSIQKVKAINLAKKSSKKRNKNAH KGEGDRVIPNLKPKHLYSGKRGIGKTDRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
539	sp_009050_gfrg	#N/A	#N/A	0.6642345	ubiquitin-specific protease 8 (UBP8); FUNCTIONS IN: ubiquitin-specific protease activity, ubiquitin thiolesterase activity; INVOLVED IN: ubiquitin-dependent protein catabolic process; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site (InterPro:IPR018200), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2 (InterPro:IPR001394); BEST Arabidopsis thaliana protein match is: ubiquitin-specific protease 5 (TAIR:AT2G40930.1); Has 10748 Blast hits to 7392 proteins in 262 species: Archae - 0; Bacteria - 6; Metazoa - 5718; Fungi - 1753; Plants - 1297; Viruses - 10; Other Eukaryotes - 1964 (source: NCBI BLink).	MPSSSSSSTLTQQRRFKFKLLYAFFSLIFSLLSRL FDSLIFSLSSAMDNLFDPDDFLSVGNDATTTSNN DDGDDDDGDLLYPQSRRRRRRLHSSPPSSSSTS TLYLRSDRLFLVCYKWWKDAREGCLGSKGEIKG DRYTTAELSDDENKRESDSDDEFVLNLKREVDD DDDNGGDDKGKGFSGREYALITESMWFTGLKW HSDYCTRIKDMGCSLAAEDNEQDVFPIQIRLYVM QQTDSLVVKISLMDNPIQLYKSAYQIFDVQSNLVC IWDFSGQMTQFFERGGALFPNGCPGQAPEEVLL ELQVYGPSFGNNAGEKKTEIAMQGSSSSGDSCG SCSIKMDGTVQSSSSPFLSLQSCTHRYGETGVL GLTGLLNLGNTCFMNSAIQCLAHTPKLIDYFLGDF KKEINYENPLGMNGELALSFGDLLRKLWAPGRM PVAPRLFKSTLENFAPQFCGYNQHDSQEFLAFLL DGLHEDLNRVKHKPYIQLNDTEGRPDVKVADEH WEIQLARNDSIIVDLYHGQFRSTLVCPICKRVSVT FDPFMYLSLPLPSTTMRTMTVTLLSTDGSMPPSP LTVTVPKSGRCKDLIQALGTAVSLQDDETLFVAE VYNHCILHSLEVPSDSLALIRDCDYLVAYRLPKES ESSLLVVFMHQRCERQFGFRMGRRSWSPFGIPL VARIPNISDGSQIHDTFKKLLRPFLNMDEDDYDDI ENSDSELESFPNSDNSWSSDDEMGDDLHVPSD FQFHITDERGMLRGRRIKVGKVLDVSVASNKVYV LVSWSKPVLRRYDTCRLSLLPEVFKHVFAAMKP QESVSLYKCLEGFLKEEPLGPEDMWYCPSCKEH RQASKKLDLWRLPEILVVHLKRFSYSRFLKNKLE TYVDFPLEELDLTGFLAHESGPRLNRYRLYALSN HYGSMGGGHYTATVYHGKGWYEFNDSFVELVR EEDVKTSAAYVLFYRRISDT
540	sp_009140_ckki	#N/A	-0.8837473	-0.937478	Polynucleotidyl transferase, ribonuclease H-like superfamily protein; FUNCTIONS IN: exonuclease activity; LOCATED IN: intracellular, chloroplast; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Exonuclease, RNase T/DNA polymerase III (InterPro:IPR013520); Has 576 Blast hits to 576 proteins in 123 species: Archae - 0; Bacteria - 44; Metazoa - 336; Fungi - 16; Plants - 80; Viruses - 0; Other Eukaryotes - 100 (source: NCBI BLink).	MQKGLEASLGCVQSKSYLYGPTYPRSFSGQQT MELKYTNGEKQANVDSNNGAVVTHDKLAVQSEY CGRPVRQESYYLWSALTPNSQKDVHNSQSHLN DGCFYPANMGNYFQTTHWLPQVPMQMVPQSYS PDYVFQDFQYFVVIDFEATCDKERNPHPQEIIEFP SVLVNSRTGQLEGCFQMYVRPTHHNLLSDFCKS LTGIQQSQVDKGVLLSDALLMHDKWLEDKGIKNT NFAVVTWSDWDCRVMLESECRFKRIRKFLYFNR WINLKIPFNKIFGSGRRNLKQAVELAGLTWEGRA HCGLDDAKNTANLLSHLMQLGFKLSITNSLKWQT ADHPSAVQVQQTVTTTPESNLTYQPQNYKTAPS VTAIPFQPANLLKEPYVYCYCYCGLTSTKRVNQNPG PKHGCVFFGCGNWTSNGGPCCQFFKWVVSAS DSGKPV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
541	sp_010180_udyc	-0.953636	-0.999057	-0.881247	BED zinc finger ;hAT family dimerisation domain; CONTAINS InterPro DOMAIN/s: HAT dimerisation (InterPro:IPR008906), Zinc finger, BED-type predicted (InterPro:IPR003656); BEST Arabidopsis thaliana protein match is: BED zinc finger ;hAT family dimerisation domain (TAIR:AT1G18560.1); Has 1233 Blast hits to 1128 proteins in 97 species: Archae - 0; Bacteria - 4; Metazoa - 232; Fungi - 219; Plants - 755; Viruses - 0; Other Eukaryotes - 23 (source: NCBI BLink).	MEEGDSTAGSKSFSFSPSPSSPEISILNPLAFQN SLPFQNSLPFQNSLPFQNPLPFHPTNTTQTQQAE MPENNFQEFGQGEGEMGPPPPRPRRLKSDVW NHFEKIIINGEQKARCLHCKHLFSGNNKNETSHLK DHLNLRCTHKHMRVDVRQKLLSVNRRQDSSVRL ENHVFNQEESRKDMAHMVILHEYPLSIVEHVEFK RFINGLNPNFKKKRRNTLKSDILKMFVNEKANLKK MFDGHAGRVAITTDMWTASHQKKGYMAVTSHFI DSQWVLRNKTLRFVYVPCPHTTNMLLKF
542	sp_010730_piyr	-0.847585	#N/A	-0.921335	LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 4 (LOS4); FUNCTIONS IN: RNA helicase activity, RNA-dependent ATPase activity, ATP-dependent helicase activity; INVOLVED IN: poly(A)+ mRNA export from nucleus, response to cold, response to heat, response to abscisic acid stimulus; LOCATED IN: nuclear envelope, plasma membrane, cytoplasm; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: eukaryotic initiation factor 4A-III (TAIR:AT3G19760.1); Has 40925 Blast hits to 40635 proteins in 3033 species: Archae - 702; Bacteria - 21081; Metazoa - 5720; Fungi - 4311; Plants - 2419; Viruses - 28; Other Eukaryotes - 6664 (source: NCBI BLink).	MADSTEATVEAPGSSSTPESSSVPESSAASESC SAPEIKPAEVKRSWGDEVDEEFTIDQLSIKESKDL DEPEDSNIKAVITGDTPYTSAKKFEDLNLSEELTK GLYVEMRFERPSKVQSVTLPMILTPPYRNLIAQA HNGSGKTTCFVLGMLSRVDPKQQAPQALCICPT RELAIQNLEVLKKMGKYTGISAERAVPENLDNPA HTPISKRPPVTSQVIIGTPGTIKKWMSFRKLSMSS MKILVFDEADHMLAEGGFKDDSLRIMNDIKRSAP SCQVLLFSATFSERVKAFVKSAIQDYNQLFVKKE ELSLQSVKQFKVRCPGELEKVQVVREKILELCQK VGQTIIFVRTRNSASYLHKALVDLGYEVTTIQGAL THDDRDKIVKEFKDGLTKVLISTDVLARGFDQSQ VNLVVNYDLPVKHDPNKKYDPNKAEPDCEVYLH RVGRAGRFGRKGAVFNLICDDIDNKVMGKIENHF GCEVIEVPDYRNDEDFEKALKDAGLL
543	sp_011570_rfcq	-0.989672	-0.931667	-0.975437	Nucleotide-diphospho-sugar transferases superfamily protein; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 2 (InterPro:IPR001173); BEST Arabidopsis thaliana protein match is: Nucleotide-diphospho-sugar transferases superfamily protein (TAIR:AT2G39630.1); Has 22705 Blast hits to 22656 proteins in 2589 species: Archae - 933; Bacteria - 16103; Metazoa - 288; Fungi - 260; Plants - 98; Viruses - 23; Other Eukaryotes - 5000 (source: NCBI BLink).	MAEKQEREKSKFSIIVPTYNERLNIALVVYLIFKHL RDIDFEIIVVDDGSPDGTQEIVKQLQQVYGDDRIIL AARPKKLGLGTAYAHGLKHASGNFIVIMDADLSH HPKYLPSFIKKQMETGASIVTGTRYVSGGGVHG WNLMRKLTSRGANVLAHTILWPGVSDLTGSFRL YRKSVVEDLIRSCVSKGYVFQMEMIVRASRKGHL IEEVPITFVDRVYGSSKLGGSEIVEYLKGLVYLLFT T

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
544	sp_011640_gwnu	-0.893331	#N/A	-0.85722	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G34310.3); Has 84 Blast hits to 78 proteins in 24 species: Archae - 2; Bacteria - 4; Metazoa - 9; Fungi - 8; Plants - 55; Viruses - 0; Other Eukaryotes - 6 (source: NCBI BLink).	MPSGAKKRKAAKKKAQKQSHNTDGSHLQGIDEA KSHDERESDGGEGERESSEERVVSKNEEGEQH ATEFSKESKSEENYDDMRVNVEDDVRKTKNGKKH KSSSSSSSSSSSSEDEVVIEKEKSVVSEVQSSL PEASRENPVISETVYASVEESIPVEESVKQVLSLP QETNTRSESADVENSMIPDDIELGMKQNLDEVLS SADGKDKVSSGMTEFLQEKIEKKVIGVEEEGPAN LSKSLHNTGIDAEASLPPSSGVTSADACCEDNHV KESETHKYTEEQPLVSSAPLPVRRTSLMNCCGLF EVLSGGASR
545	sp_012380_tgua	-0.9817327	-0.9823167	-0.9944449	Single-stranded nucleic acid binding R3H protein; FUNCTIONS IN: nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Single-stranded nucleic acid binding R3H (InterPro:IPR001374); BEST Arabidopsis thaliana protein match is: Single-stranded nucleic acid binding R3H protein (TAIR:AT2G40960.1); Has 389 Blast hits to 389 proteins in 86 species: Archae - 0; Bacteria - 0; Metazoa - 196; Fungi - 53; Plants - 134; Viruses - 0; Other Eukaryotes - 6 (source: NCBI BLink).	MDGGNAVEDPVAPESWEVADVEESMKRLLLSA SSSAKDSATPSSSSSNCIAADDLTASGPLNLAFS SGAENAGVSGLSDDAINQVDQFLREALQNPRER LSILRLEQDVEQFIQDPTQQQLEFQQLPTSYLRLA AHRVAQHYFLQSMVLLDNSLPDGSGSRIIVQKTS TECRFPRIRLADIPVSLPQEDSGVVKVAIKQRPNK HSSSSSVGNSGSQKANSKSVEERKEEYNRARAR IFNSNSQNGSPGLKPEIEAKKAENLQLPSSGPPK VDDKPASGGSDLSSGRAVSVVDSTSGSSRSARS RTEKDPISRYRPSNRVAIFRDRDIDRKDPDYDRN YDRYLQRFDPGFGFHAGPYTVQPMYSPAVNYNT EFPQLGSTPRPQISADPPRPLPQHLPLWASPSN PAGLGYGHPETMMAAYSPNHVNPPRSTSAIYLHT TQYPCQRPGMPFIHPHEHIHQSYQQSQQQQHD ATFGFARPR
546	sp_012530_qgin	-0.9194451	-0.9858884	-0.9831438	LOCATED IN: mitochondrion; EXPRESSED IN: shoot apex, embryo, flower, seed; EXPRESSED DURING: petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; BEST Arabidopsis thaliana protein match is: PIF / Ping-Pong family of plant transposases (TAIR:AT3G55350.1).	MAASPKPPSTKRRTTKRKTTATTATTLSPLTAAI TTALTSSHSFFHNHELHLLLPSQSVSLQSSLSSVF LSLSKLQPHFQLALPLSSPSSSSCCWFQRFLLSSS NDEDPRWVDFFRMSKPTFYRLNSILSNSISSVLP SSISPDFAIGAAICRLAHNWSFSFVGKRFGLSSSD ACRVFYAVCKVICEQLGGYLGGFGSELRDIVGF ERISLINCCGVLGVEKFPIEGELLGGNGFLMIQAL VDSDGRFLDVSAGWPSTFDPDSILRQTRLFLSVD ESKELLNGPEVELIDGSLMPQYIMGDSCYPCLPW LLTPFARRGEAEDSSSSKKEFNEAHKRATDLVR MAFGRVRANWKLLDRKWKEEFAEFLPFIITACVL NNFLMKHGELLSHGDLGCVREQDIPVFDGEVDE SAARIRDLLALHLSRVTQDN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
547	sp_012600_uczs	#N/A	#N/A	0.854913	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family; CONTAINS InterPro DOMAIN/s: Late embryogenesis abundant protein, group 2 (InterPro:IPR004864); BEST Arabidopsis thaliana protein match is: hydroxyproline-rich glycoprotein family protein (TAIR:AT3G52460.1); Has 868 Blast hits to 868 proteins in 31 species: Archae - 2; Bacteria - 0; Metazoa - 2; Fungi - 0; Plants - 862; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MTDPSRPATGYPAPPQNGYPPPPPPQQSNYYY PNHNPNPYNYNAPQAYYPAAPPRATFFRRFFAA LIALFIIFGVILLIVWLVLRPRIPSFTVDSVALSSLNV SSSPSQISGTWSLHFSVSNPNHKLRVYYDEIEAR LYYRSSLLTRNQLAPFDQDTKNITTMDAKLVASS AYIESELVNRIKGDRSTGMVNFDVQVLALVRFKA GGWRARRRILRVFCDNLSVSFSSNSTMGTLSGG PKRCSVGI
548	sp_012680_xqaa	-0.86227	-0.865343	-0.844037	Intron maturase, type II family protein; FUNCTIONS IN: RNA binding, RNA-directed DNA polymerase activity; INVOLVED IN: RNA-dependent DNA replication, RNA splicing; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Intron maturase, type II (InterPro:IPR000442), RNA-directed DNA polymerase (reverse transcriptase) (InterPro:IPR000477); BEST Arabidopsis thaliana protein match is: Intron maturase, type II family protein (TAIR:AT1G30010.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MVLIRLIRITGSGILPLLRHSYRLNGFALSRSFSGF HNNNWRKPDPDDPLSLMKEDGVAVCSHMWIEN FREPEKSASNLTSYLRKFELWVLAYQKVVADET GSYLPRHSIQKPALEDLLSLRNAVLDDRFKWGAR LNFVIRSPRDKTEYESLSKRKIKTILTTTQPAPYQD RIVQEVLFMVLEPVYESRFSQKSYAFRPGRTAHT VLRVIRRSFAGYLWYLKGDLSTILDGMKVGLVINA LMRDVRDKKVIDLIRSALVMPVIETKIDDGEKQKK KKRKYQKKKVLADDEPKPDPYWLDSFFGFAPEE AEKVPCWGHCGILSPLLANIILDELDHWMEGKIK EFYHFSKSDVIWNNSPEGEAEQGNTSWPEFVPTS GPDKTRKMDYIRYGGHILIGIRGPPRADAAVLRKQL VEFCDQKYLLKLDNDSLPIEHITKGIMFLDHVLCR RVVYPTLRYTATGGKIISEKGVGTLLSVTASLKQC IRQFRKMNLLKGDREPDPQPCFRMFHATQAHTN SQMNKLLCTMVEWYRFADNRKKIVNFCSYIIRGS LAKLYAAKYKLRSFAKVYKISDRTLSRPLKEKKGL SPEYQNMLRMGLAESIEGLKYTRMSLVPETDYTP FPGNWRPDHEKALLEFLSLDDPKTLEDQRNCIKE QGLISPQDYISMLVWNYKRNASLSDQLSLLNSGG KNIQSCDQLEMESSNEDCVERSEDEEEDKERIN MAHV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
549	sp_013670_zccf	#N/A	#N/A	-0.8602336	PHD finger family protein; FUNCTIONS IN: DNA binding, zinc ion binding; INVOLVED IN: regulation of transcription, DNA-dependent; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: DDT domain (InterPro:IPR004022), Zinc finger, PHD-type, conserved site (InterPro:IPR019786), EF-Hand 1, calcium-binding site (InterPro:IPR018247), Zinc finger, PHD-type (InterPro:IPR001965), DDT domain superfamily (InterPro:IPR018501), DDT domain, subgroup (InterPro:IPR018500), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: metalloendopeptidases;zinc ion binding;DNA binding (TAIR:AT5G35210.2); Has 3640 Blast hits to 3265 proteins in 204 species: Archae - 0; Bacteria - 0; Metazoa - 2369; Fungi - 398; Plants - 598; Viruses - 0; Other Eukaryotes - 275 (source: NCBI BLink).	MESSGGRSRGRPRKRRRGDEQNGENVDGEGE VQDMVRDDNNDDDNEGEDVKNAGNSRKRGRPP KRRAVEIRGDSVVGRYVLKDFEGSGVFLGKIVSY DSGLYRVNYEDGDCEDLDSDELREYLIGEERQKF VGDLLRRKQKLDENVADLLEKKQKSDENVSPKT VVKSKSERKKRNVRKVVDEVEKGNVAEPIDIMDK CDCGNMLNDTEKVEVPISSGRFDEEVDGSEVDD DDVSDSSSDSCEYGQGWETRSEIEIPAIPPPELP CSSGSIGVPEEHVPHLFSVYGFLRSFSIRLFLSPF TLDDLVGCLNCTVQNTLLDAIHVGLLRALRWHLE AQATDGSELASKCLRCIDWGLLDILTWPVYLVHY LMVMGHLNRKEWKRFYIDVLQKDYCTLSIGTKLM VLQILCDDALDAAETRAEIDNREESEFGVDSDGM SSFSLENGPRKVYPRSAKTSIAKERENLILEAKNN TPFGSDSLTVKGKNQDTKSDNIDEDENGDECRL CGMDGVLLCCDGCPSAYHSRCIGVNKLSIPTGD WFCPECKISKLGPPITMKTSLRGAEVFGVDPYAQ LFLGTCNHLLVLKVSLKTSPFVRYYNCNDIPQVVT ALCSSEYHAVSYSGICQAILKYWELPQVLLPPPIS CQTAMVLEKKDDDSVLLSAETGNQVLNKVEPEN YASGVGESSLNYVSGPGNEKSSSTHGYDDVLIG APVQNDIAVPRMNGMNESSTMSNKLSKNAGLSG SSPQGLADGSSVMGISSCTSGYKSSSAGQRNSV CLPGNVYLPGKYRNEAGVRAVQQWKDNGNCLYI GTSFKTHGYINHYVHGNFAATAAANFAGLSSDEN RLSEVHASDPRKVMSAKVTLQAKAFSSATSRFF WPSTEKKLVEIPRERCSWCLHCKAPVSSRKACL LNQAALSATRAAMKFCSGLKLVKNVEGGLYGITA YTLYLEESMRGLIVGPFHSASYQKQWRKKVEQA STCREMKSLLLELEENLCINALSSEWTKLVDSLM DELSVPQSACTAVSNQKRGPVKRKKHLATAERIV EDSNNSLNDLNWRSGKLSKLVFQRATLPCPVLR
550	sp_013700_jjuf	#N/A	-0.8133394	-0.8597794	uncharacterized protein	MSPDVLSSSSTQDIGKSPKDNKKMYTKKKDPLA SLNKSKCEKSEKLGALNQLLRFCILKKRE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
551	sp_013730_tmpz	#N/A	-0.9073385	-0.8577035	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: transferase activity, transferring hexosyl groups, UDP-glycosyltransferase activity; INVOLVED IN: metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: 4 anthesis, LP.02 two leaves visible, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: UDP-glucuronosyl/UDP-glucosyltransferase (InterPro:IPR002213); BEST Arabidopsis thaliana protein match is: UDP-glucosyl transferase 76C1 (TAIR:AT5G05870.1); Has 7583 Blast hits to 7530 proteins in 440 species: Archae - 0; Bacteria - 352; Metazoa - 2017; Fungi - 25; Plants - 5047; Viruses - 86; Other Eukaryotes - 56 (source: NCBI BLink).	MGKLAIATATATATASMEKQVIKLGGRVLLFPYPY VGHTTPMFNLGNALYSKGFSITVIQTRLNSSDQT TFPNFTFHYIEEALWSQEALASVPDPWEAVTLLN KSFPLPFRDSLSQILEDASKDIEPVACLISDPLWD FAGTIADELNLPRLALRTGGLLAFVLYDSVPLLRQ KGYYPIQESEQEEAVSEIPPLKVKDLPPEAIHDTL AVVVKEAKTYRLGYIGNTFKELEGSFLDLFGRSLP GTPIFPIGPLHKYPSTYMGEKLDEEHSSITWLNTQ APNSVLYVSFGTIAAISKEQFLEVAWGLAKSQQP FLWVVRPKMINGLVKDDNDIFPNGFLETVSGRGY TVTWAPQLEVLAHPSVGGFWTHCGWNSTIESIC QGVPMICLPFFGDQKMNARHVSDILRIGLHLEKGI ERQNIERAIRRLMLGKEGEEMRERVAALKEEAQS CLMEGGSSFKALDQLTNHILSFCSAEIPKKHRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
553	sp_013960_ecsx	-0.850635	-0.917023	-0.942959	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity, nucleic acid binding, ATP binding, ATP-dependent helicase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), RNA helicase, ATP-dependent, DEAD-box, conserved site (InterPro:IPR000629), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEA(D/H)-box RNA helicase family protein (TAIR:AT5G60990.1); Has 60083 Blast hits to 51461 proteins in 3186 species: Archae - 589; Bacteria - 20345; Metazoa - 13239; Fungi - 7387; Plants - 3446; Viruses - 273; Other Eukaryotes - 14804 (source: NCBI BLink).	MFHHSPPVVGGASSNHLSFSDTDDIFGFVSLEEI DESVYNLAIPESREDNVKEKSGTSHKAKTRKRVD TIGAECDDVGIEVDKVLEQGSVKEKKNKKKRRKT KQVQTQPKDGDVEIEADKEMEEGYVNQKKKRKL KQVQTQVTETNTEDGEQLVDENEFNAWHELRLH PDLMKAIYKLGFKQPTPIQKACIPAAAHQGKDVIG AAETGSGKTLAFGFPILQRLLDEREKGEKLMSED AEAVEILAPRGFIRALIITPTRELALQVTDHLKAAA TSTNFRVVAIVGGMSTEKQERLLKERPEIVVGTP GRLWELMSGGEKHLVELHSLSFFVLDEADRMIE NGHFQELRSIIDLLPGTSGVTEKHFQDVQNCTTIS NFQRKKRQTFVFSATIALSRDFRKKLKRGSLKSK HSATDGLSSLEALSERAGMKPDAAVFDLTSESIM TSNLEESFLECSEEDKDAYLYYILCVHGQGQTIVF CSSITALRHVAALLRTLNISVCVLHGEMQQRARL KALDRFRENKNGILVASDGAARGLDIPGVRTVIHY QLPHAAEVYVHRSGRTARGNTDGCSLLLVSSNE RSKFASLCKSFGKDTLQRFPIVNSYMPEVRKRLS LACQIDKISRRDSKEKANKSWFKRNAESIGLDVE DNDSEEERVNKCRQKKTSSLQMKKLQQELNTLL SRPLQPKSFSHRFLAGAGVSPLLQHQLEALAQQ KCINDKSSGNIKKNVVVIGQDCVEPLQALRRAGP EVRLDLNALAEKQQDMEKLRRKRKDAKKRQKEQ KRQEKKRRKEAIGS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
554	sp_014140_gqgg	-0.336774	-0.961377	-0.964112	Outer arm dynein light chain 1 protein; INVOLVED IN: biological_process unknown; LOCATED IN: vacuole; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat, typical subtype (InterPro:IPR003591), Leucine-rich repeat (InterPro:IPR001611); BEST Arabidopsis thaliana protein match is: Outer arm dynein light chain 1 protein (TAIR:AT4G03260.2); Has 24271 Blast hits to 14439 proteins in 747 species: Archae - 6; Bacteria - 9514; Metazoa - 8699; Fungi - 1002; Plants - 3127; Viruses - 44; Other Eukaryotes - 1879 (source: NCBI BLink).	MAVLFKCFSKFIGRKKKVKDAIVSDNGVRTLQVK VEHPMKPSENVVDDLKSTSFSVPVPFDIPEISTCE VKLMNSDSPVEGDQEEVAYEGEDEHDETSLMSR DNSDSNLQARDNIGTPRNLSFHQIGSSPCAAEED DVAKNNEEIIDTGRNGHVSDPGMAREEKWASPR LSRSCSNIESREVLKTTQSFKQLQELSNTNENEA RSPCCSHSSALTHRSADKVMLKKHSYSQILPSRS KKLWWKLFLWSHRNLQKPEIASKTRVSSLLLASV NQQGGYSSDTMEPTRVKQLKKVESTGSFSEQC SSYNKSSHEEDNNNGLEGRFSGLWPQNQWVAF SGESSSLKRVDDWVKDLEFQPAISALEEGNVVD DDDEQIVFPSSPKVDDSVVKNPTHFPRHNSKINL TEELLHANSVIQSLNSSSTVAHVSGMGLKVIPIISH FSSLRSINLSNNSVQITQGSLPKGLHVLDLSRNKI TTIEGLRELTRLRVLDLSYNRISRIGHGLPSCSLLK ELYLAGNKISEVEGLHRLLKLTVLDLSFNKITTTKA LGQLVANYNSLLALNILGNPILSNLSSDQLRRAVC GLLSKLAYLNKQPVNQQKARFVSMDIIAKSSSRS NSWNARRKSVKKTGQGGSIMLSGSVRSNASVRQ RTRNKAKREGNQQSRLKIRTTTTTTS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
555	sp_014440_dsja	#N/A	#N/A	-0.870598	Cation efflux family protein; FUNCTIONS IN: cation transmembrane transporter activity, efflux transmembrane transporter activity; INVOLVED IN: cation transport, transmembrane transport; LOCATED IN: membrane; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Cation efflux protein (InterPro:IPR002524); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G38380.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MEDSHHHHHRPQRLSVPPRTNVLTALRHPSAYP LTPTPTPSKTRLSRSTSTINDKSSIPFLFLLLFSLR SLYSVLPFLRSSPSFSLFPFSFLVSLLSFLLTLSFS LYSKHPRLPLFSLSSLSQSQVRHLLFKSLVLGMV FLLRFQALRYCGTAAMILAELSGNMAFRFWKDQ NWDRDGHSRVRGFFALFSGLFLLSVSWDRMDC FPLSYSNLGKIQNSLMVVVDGNCLRIVPMLLPFLS GFLGCYERDLMNWGTIRQLGRKRVQLISLFFTTS MLFFPAIINMLVFEAEGGSVSIMNLGWSLANTVLF GVLLNEWYNDEKRVNPSDSEKEFLITFLCTLVLE LVYFPELSLWGLLICGFLLWVAIQQLNWARSRYV ELGAESSETFLSMVMRPIRHVLSERKSRKIALFLM INTGYMVVEFVAGFMSNSLGLISDACHMLFDCAA LAIGLYASYISRLPANSQFNYGRGRFEVLSGYVN AVLLVLVGALIVLESLERILDPQEISTNSLLAVSIGG LLVNVVGLIFFHEEHHHAHGGSCSHSHSDSHSHT HHDHDHVTEVMHHDGVLVGESKNGSHDMHGES HQHDHVHSHGDHDHSHTHGHHDHSHHHGNHD HDHSHGHHDHDHSHAHHDHDHSHAHHDHSHAH HDHSHSHDQNHSCGHHDQHHEQGHHIHHDHND NEVCKSIPTVPGVDSCSHSHHLDSKTESSHSKKP RHQHIDHNMEGIFLHVLADTLGSVGVVISTLLIKY KGWLVADPACSIFISVLIVSSVIPLLRNSAEVLLQR VPRAHEQDLKAAVTDVMKIKGVSGIQNLHVWSLT NTDVISTLHLRISSEIDQAATKARVMHVLEDAGVK DLTLQVEGVKSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
556	sp_014450_yfzt	-0.8817235	-0.813539	#N/A	eukaryotic translation initiation factor 2 (eIF-2) family protein; FUNCTIONS IN: translation factor activity, nucleic acid binding, GTP binding, GTPase activity; LOCATED IN: plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Small GTP-binding protein (InterPro:IPR005225), Translation elongation factor EFTu/EF1A, domain 2 (InterPro:IPR004161), Translation initiation factor 2 related (InterPro:IPR015760), Protein synthesis factor, GTP-binding (InterPro:IPR000795), Translation elongation/initiation factor/Ribosomal, beta-barrel (InterPro:IPR009000); BEST Arabidopsis thaliana protein match is: eukaryotic translation initiation factor 2 (eIF-2) family protein (TAIR:AT1G76720.1); Has 208612 Blast hits to 135501 proteins in 4311 species: Archae - 1435; Bacteria - 54544; Metazoa - 70572; Fungi - 21896; Plants - 9806; Viruses - 724; Other Eukaryotes - 49635 (source: NCBI BLink).	MGRKKNCVDDDETSVATELSEESGVTEEDKVVP TGGGKKKGKKGNSKGAQSRDEEDEEVPEVAFA GKKKSSKKGGVKAGFELLGDEDNDDQVKDGSD DEEDVSVVSFTGKKKSSKGSKKGGGKASFALLG DDGDEDDEVKGQSEGEDDVPAVSFTGKKKPSK GSKKGGSKTSFALLDDEDNDGDENVQSEEDEAP SMSFVGKKKPSKGSKKGSSLFSAALLEEEDGEE DDSHSEAPSVSKTAYDEVDEADVVVAFTGKKKS SKKKGSSALTALGDDITTKMETKVIANNAEVDDYL PQEKVSAVVPETSKNKKKKKKGGRTAEEQDDLD KILAELGEGPAVSKPEPTPAPSQVEKADTQLEPV GLTEAAGEKEGEDEGIVESAAAKKKKKKKEKEKE KKAAAAAASSAEIKVEKEDVKNETKNKAVDKKLP KHVREMQEALARRKEMEERQKQEEEERLRKEE EERLRQEELLRQAEEAKLKKKEKEKEKLRLKKEA GLFLTAKQKEEQRKREIMKNQFLATAGNKLPASD NGAPTKRPIYTTKKSKSNRQTNGVKAQTLEDME TNQEAVGDSELVETEKVDEVCTIDVEEKSEVADE IEDVVAEEDNEDDDEDWDAKSWDESNLSLPATR AFEDEETDSGDQSLVKKEVKQAAQSVRDTGRPA ATANEAHSSQATKGQTSLSKQKHSDSEAVGKKR SKETEVQKKKPSSDTQAKNGEGDLRSPICCIMG HVDTGKTKLLDCIRGTNVQEGEAGGITQQIGATY FPAENIRERTRELKADAKLKVPGLLVIDTPGHESF TNLRSRGSGLCDIAILVVDIMHGLEPQTIESLNLLR MRNTEFIVALNKVDRLYGWKTCRNAPIVKAMKQ QSKDVQNEFNMRTIGVITQFKEQGLNTELYYKNK DMGETYNIVPTSAISGEGIPDMLLVLVQWSQKTM IEKLTFRDEIECTVLEVKVVEGHGTTIDVVLVNGV LHEGDQIVVCGMQGPICATIRALLTPHPMKELRV KGTYLHHKQIKAAQGIKISAQGLEHAIPGTALYVV GPNDDVEEIKKLAMEDMESVLSRIDKSGEGVCV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
557	sp_015320_cmxs	-0.873832	-0.855378	-0.918379	uncharacterized protein	MDDSYAAAADDGSGWFEVKKKRRSSSKFSIQ NWVGGFSRQQVNGKGRSSNVKYRSNISKPGKD SVQRGVSSHSSSSKEHEKDTCLDKCVVGDDTG RPKTPPLKIITTSSQSSSTANENVTEKKSDIVQRI KWGDLEDDALDCGTAAVSEIRFGEIGDDDLAMC CTSRNVCDLSACNLSKDYPVVNNSTFEVLEEASD MCYSKPPPFVDESLEENCKEVIEVASTDVGMLST TTKVIDANCSGLDHKELVNNDVENHNEDHACGS SNDVFGCSSIEEPAKIVELSAPILTPGVCDLGFSS VKDEDNADGMLIARDMDISPSVPNLAEAPENCNA ASVPADCGSGIPNSSSTDNLLNTPTMNINREPEG GESKERFRERLWCFLFENLNRAVDELYLLCELEC DLGQVKEAMVVLQEAESDFRELNIRVEEFETLKT SCPPTSGAVMSIKGDHRRPHALSWEVRRMTTSP HRAEILSSSLEAFKKVQEEREKERASGEVKKRHK RLPTMDEHKRENIPDTSELETKSRRKSGISDAAR LNTSRERRSTEAGRSSKAPVQNSRVALVTSASD PTTTKLPSRGSSAAFVGGKGKKEQLGSLAEMEK LFEGDVFRRHNVAGDKEKEKRNSQPWRPVDAW KEKRNWEEILAPPRRLSSRVSHSPCMSRRSIERA RVLHDKLMSPEKKKKTAVDMKKEAAEKHARAMR IRSELENERVQKLQRTSEKLNRVSEWHAVRSMK LREGMYARHQRSESRHEAHIAQVVKRAGDESIK VNEVRFITSLNEENKKFLLRQKLHDSEMRRAEKF QVMKTKQKEDMAREEAVIERRKLIEAEKLQRLAE TQRRKEEAQLRREEERKASSAAREARTMEQLRR KEERAKAQQEEAELLAQRLAERLSESDQRRKVY LEQIRERASMDFRDQSSPLMRRSLNKEGQGRLT PPRNSDDNQTNNPSSESGPVNNGIGMQQSLKR RIKKIRQKLMALKHDLPENPAAENVGIGYRTAVAT ARAKIGRWLQDLQRLRQERKEGAASIGLITAEMI KYLDGKELELQSSRQAGLLDFISSALPASHISKPE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
558	sp_015800_tqrp	-0.3514833	-0.907113	-0.359334	multidrug resistance-associated protein 14 (MRP14); FUNCTIONS IN: ATPase activity, coupled to transmembrane movement of substances; INVOLVED IN: transport, transmembrane transport; LOCATED IN: plasma membrane; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: ATPase, AAA+ type, core (InterPro:IPR003593), ABC transporter-like (InterPro:IPR003439), ABC transporter integral membrane type 1 (InterPro:IPR017940), ABC transporter, transmembrane domain, type 1 (InterPro:IPR011527), ABC transporter, transmembrane domain (InterPro:IPR001140), ABC transporter, conserved site (InterPro:IPR017871); BEST Arabidopsis thaliana protein match is: multidrug resistance-associated protein 6 (TAIR:AT3G21250.2); Has 673798 Blast hits to 360971 proteins in 4075 species: Archae - 12279; Bacteria - 544381; Metazoa - 13093; Fungi - 8967; Plants - 6941; Viruses - 21; Other Eukaryotes - 88116 (source: NCBI BLink).	MLSSFELFANSHCDNVNCGDRSVCVHFMLVSEG KIVRASYHELLMTSEAFQNLVNKYDTTASQNSKE VLRDTAKTVTSKILQEDTKENLNVSDQLIKREERE MGDAGIQPHLQYLKYGNAFLYFPLSNFMQWLQH LSSPSCWNRFSGLQRLSMIQHLWEEYSVGLPLV YLKNCCIIYYSDHTLAIKRIGYTLGIMAGSNCSASL KDTQHSVPGGTLASQLTTHNSQPLVCI
559	sp_016290_hnar	#N/A	-0.8473838	#N/A	minichromosome maintenance 8 (MCM8); FUNCTIONS IN: nucleoside-triphosphatase activity, DNA-dependent ATPase activity, DNA binding, nucleotide binding, ATP binding; INVOLVED IN: DNA-dependent DNA replication initiation, DNA replication; LOCATED IN: cellular_component unknown; EXPRESSED IN: shoot apex, embryo, flower, seed; EXPRESSED DURING: F mature embryo stage, petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), Nucleic acid-binding, OB-fold (InterPro:IPR012340), ATPase, AAA+ type, core (InterPro:IPR003593), DNA-dependent ATPase MCM (InterPro:IPR001208); BEST Arabidopsis thaliana protein match is: Minichromosome maintenance (MCM2/3/5) family protein (TAIR:AT2G16440.1); Has 4027 Blast hits to 3851 proteins in 539 species: Archae - 371; Bacteria - 134; Metazoa - 1249; Fungi - 930; Plants - 423; Viruses - 3; Other Eukaryotes - 917 (source: NCBI BLink).	MDGGGGESIVMTMMQRPSAAEPCCDLLHIWPIY FPQIELSFQDWRYILISKLVCFFATPRGNHLVSQV KHEDDTFSLQLNFQEFRKICDLECFYYNLEDYPK EALLCMGAAIHKVTSFQEKSKLESHGKINIRLHNY PESLIALRNLKAAYIDRLVSVRGTAVKVSTVRPLV MQMTFKCIQCGADIIRFFPDGKYSPPATCIMRCC KGKTFVPIRPTAQSIDFQKIRLQELLKPENHEEGR VPRTVECELTEDLVDVCIPGDVVTVTGIIRQINNY MDLGGGKSKSKHQGLYYLYLEAVSVTNSKSLSIP EDSEYTTADARATEMSDLFSFSPRDLEFIVKFNE EHGPDVFRQMIQSICPSIYGHELVKAGITLALFGG VRKHSMDQNKVPVRGDIHVIVGDPGLGKSQLLQ AAAAVSPRGIYVCGNATTNAGLTVAVVKDRMTS DYAFEAGAVVLADRGLCCIDEFDKMSAEHQALLE AMEQQCVSVAKAGLVASLSARTSILAAANPVGG HYNRAKTVNENLKMSSPLLSRFDLVFILLDRPDE HMDKRLSEHIMARKQPDGQGVFILGSPDILDSLL EGHLSEDRVTLT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
560	sp_016390_trim	-0.871232	#N/A	-0.89559	related to ubiquitin 1 (RUB1); INVOLVED IN: protein modification process, response to auxin stimulus, protein neddylation, ethylene biosynthetic process, embryo development; LOCATED IN: plasma membrane; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Ubiquitin subgroup (InterPro:IPR019956), Ubiquitin conserved site (InterPro:IPR019954), Ubiquitin (InterPro:IPR000626), Ubiquitin supergroup (InterPro:IPR019955); BEST Arabidopsis thaliana protein match is: ubiquitin 7 (TAIR:AT2G35635.1); Has 16064 Blast hits to 7349 proteins in 721 species: Archae - 0; Bacteria - 36; Metazoa - 7330; Fungi - 1879; Plants - 3570; Viruses - 331; Other Eukaryotes - 2918 (source: NCBI BLink).	MQIFVKTLTGKTITLEVESSDTIDNVKAKIQDKEGI PPDQQRLIFAGKQLEDGRTLADYNIQKESTLHLV LRLRGGTMIKVKTLTGKEIEIDIEPNDTIDRIKERVE EKEGIPPVQQRLIYAGKQLADDKTAKDYNIEGGS VLHLVLALRGGACSASA
561	sp_016410_mfuo	-0.973543	-0.988487	-0.987824	DGD1 SUPPRESSOR 1 (DGS1); INVOLVED IN: galactolipid biosynthetic process, hydrogen peroxide biosynthetic process; LOCATED IN: mitochondrial outer membrane, mitochondrion, protein complex; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Nuclear control of ATP synthase 2 (InterPro:IPR013946); Has 223 Blast hits to 223 proteins in 116 species: Archae - 0; Bacteria - 2; Metazoa - 1; Fungi - 125; Plants - 59; Viruses - 0; Other Eukaryotes - 36 (source: NCBI BLink).	MEVASPSQEDAESRNFRTLTLLYSNYIWTRFSNS PFIGKISSLCRQFGRAKSRRQKLCLPLPLPSKSVD SSLILTEASSIYDVLEDVFEHILSNMHNVQKNLRY WQAKAEGSNAQKAYFMIFERGPRALFDETVEFV RNFIAEDSSFQILCHSASGFISERVAVLDSLQSAL ATFLAQVYMMIDKQGEDLVTDPQKALPSLLVTIN GLFSDLENTIAQVHASMSLIDSCAGENKSSALVF NKMTEVQLEQPQWSDSEIRDAYDIINQNIQKLDS YISVLVARHQKPRRMTRYWIRYTCGAVGLSVCSL WIVRHSRLVGSPDINWFSDAKESTVSFWTEHV EHPILSIRDELFETFRKRHRGVMEAEEVHLTANSL RRMLVAFCEQANGEKLPENATDLDLLQIVMGRY EKEMMHPVKSMFCGDLARAMLIQVQKLKLDIETA MLELNQLKANEINFAILAALPAFAVIVLSMILVRS WIMQDKGAEGRGRLARRQRRLLIVEIEKPRIMHYQ SCMDMGQAQNAQCSFGLVLYTLDRLYRAVERH AKMTGEWSCLRQDIIDLAKPDLQITYKLTITSRME RVYDCLLPSLKQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
562	sp_016450_fxnk	-0.8507253	#N/A	#N/A	molybdopterin biosynthesis MoaE family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: Mo-molybdopterin cofactor biosynthetic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Molybdopterin biosynthesis MoaE (InterPro:IPR003448); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSEEDKTLIEILEENNPIDIIKYINFVRAPKCGAIAT FSGTTRDTFDGKAVVELRYEAYVPMAIRCIKSICT SARETWNLNLIAVAHCLGPVPVGGTSVFIATSSV HRVDALDACQYLIDELKATVPIWKKEVYANGEVW KENSEFLERRTEIGPVPVGETSVFIATSSVHRVDA LDACQYLIDELKATVPIWKKEVYANGEVWKENSE FLERRSELSENFRPRPLNRSSEKKGCCGTKVRV DEERSNKKSNENTGINQ
563	sp_016990_fuiw	-0.8744432	#N/A	#N/A	YODA (YDA); FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: stomatal complex morphogenesis, embryo development ending in seed dormancy; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: mitogen-activated protein kinase kinase kinase 3 (TAIR:AT1G53570.1); Has 133707 Blast hits to 131404 proteins in 4998 species: Archae - 148; Bacteria - 15191; Metazoa - 50613; Fungi - 12908; Plants - 32496; Viruses - 572; Other Eukaryotes - 21779 (source: NCBI BLink).	MPSWWGKSSSKDAKKKTGDETILDAIRRKFRNS SDVKLDSRSGASRRHCIDTVSERGAQSRAESRS PSPSKVSKHVGRCQSFAERPHAQPLPLPGLHPA NVSRTDSGISLSSKPRTEKSSKKSSLFLPLPRPSCI RSRADPAELDSDVVTNSVSSASSIESDDLTESRL RSPQAFDLDIGARTAAGSPSSVVLKDQATKAAQ NDYVEAQKPVSLPLSNQSTTASPRRRRRPLGNHV PNLLVPPHGAFCSAPDSSIMSSPSRSPLRTPNSE QILNAAFWASKPYPDVTLLGSGHCSSPGSGQNS GHNSMGGDISGQLFWQQPSRGSPEYSPIPSPR MTSPGPSSRVQSGAVTPIHPRAAGSGSDMHAS WGDDGKQQSHRLPLPPVTIPSPSPFSHGNSAAT SPSVPRSPARAENLPSPTSRWKKGRLLGRGTFG HVYLGFNSETGEMCAMKEVTLFSDDPKSKESAK QLQQEIALLSRLRHPNIVQYYGSETVGDRLYIYLE FVSGGSIYKLLQEYGQLGELAIRSYTQQILSGLAF LHGKNTVHRDIKGANLLVDPNGRVKLADFGMAK HITGQSCPLSFKGSPYWMAPEVIKNSNGCNLAV DIWSLGCTVLEMATTKPPWSQYEGVAAMFKIGN SKELPAIPDHLSDEGKSFVKLCLQRDPQQRLSAA QLLEHPFVKNTAPLVRPIPSPEPADTPIPINILNGT RSLGVGVERNLSSLDSERLAVHSSRVSPARPRS SEYYIPRNISCPVSPIGSPLLNPRSPQHLNGMMS PSPISSPRTTSGSSTPLTGVSGAIPFHHSQSQQS LYMQEGFGTLPKSPNSVFQNAPCYHDPNPNILR GMQSGSGPHAITLVSPKNGNHVKQVERERSCQ GGDPCNGRTILVDRVSKQLLREHVKQQNPSVEF SPRSTLPSRTHHGL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
564	sp_017400_qqpq	-0.884304	-0.963398	-0.993378	general regulatory factor 9 (GRF9); FUNCTIONS IN: protein phosphorylated amino acid binding, calcium ion binding; LOCATED IN: chloroplast stroma, nucleus, plasma membrane, cytoplasm; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: 14-3-3 protein (InterPro:IPR000308); BEST Arabidopsis thaliana protein match is: general regulatory factor 12 (TAIR:AT1G26480.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MASTKERESFVYIAKLAEQAERYDEMVEAMKNLA KLDVELTVEERNLLSVGYKNVVGARRASWRILSS IEQKEEGKGNEQNVKRIKEYRQKVELELTNICSDI MTVINEHLIPSATAGESTVFYYKMKGDYYRYLAE FKFGNERKEAADQSMKAYESATTTAEAELPPTH PIRLGLALNFSVFYYEILNSPERACHLAKQAFDEAI SELDSLNEDSYKDSTLIMQLLRDNLTLWTSDIPEE GEDFQKADGAAKALAGDDAE
565	sp_018230_yayi	#N/A	-0.915682	-0.983539	GNOM (GN); FUNCTIONS IN: protein homodimerization activity, GTP:GDP antiporter activity; INVOLVED IN: in 13 processes; LOCATED IN: intracellular; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: SEC7-like (InterPro:IPR000904); BEST Arabidopsis thaliana protein match is: GNOM-like 1 (TAIR:AT5G39500.1).	MTSLELQPASFPSMVNTEIGAVCLAVMRHNVRWG TRFTFADDQQDHSLILSLKTLRKKLFTSHAHWHDI DPVTYLQPFLSVIQSEETGAPITSVALSSVYYKILTL DIFDMETINIQDTMHLIVDAVTSCHFEVTDPASEE VVLMKILQVLLACVKCKAAVALSNQHVCNIVDTCF RVVQQTSAKGELLQRIARHSMHELVRSIFSRLPEI LCHELDSGGNLINHEIQQVNSDPIFVAPQPDSKD NTPTDSKLTETCSTSSSSAPFVDCPSNGTTTMDT VQYESHLMNQQFGVPCMVEILHFLCSLLNVKEHA DMNRKSNSLAFDEDVPLFALGLINSAIELGGSSF QRHPRLLKLIQDNLFCNLMQFGIMSMSPLILSMVC SIVLNLYQHLRTKLKLQLEAFFSSVILSLAQSRYG ASYQQQEVAMEALVDFCRQRTFIMEMYANYDCD LTSRNIFEELVNLLSKSAFPVNTPLSTMHVLALEG LVAVIQGMAERTADERVDLGNTSITLEEYKPFWL FKCNGYEEPGIHWVPFLRQTKYIKQRLMAGADHF NRDPKEGLEFLQSTHLLPEKLDAQSVACFLRYTT GLDKNLVGDFLGNHEEFYVQVLNEFAQTFDFKG MNMDTALRIFLETFRLPGEAQKIQRVLEAFSARY YEQSPHILANKDAALLLSYSIILLNTDQHNSQVKK KMTEEDFIRNNRLINDGNDLPRELLSELYSSICEN EIRITPEQGSGSCEIRLSRWIDLMQRSQNTSPYIA AAYKPYLDDDMFAIVAGPAVAAISVIFDHAEDEDV YQLCIDGFSAASKIAAYLHHDSVLDDIVISLCKFTT LLNPSYLEEFLLAYSNDTRARISTTTVFTIANRYG DCIGTGWQNIINCILRFYEFGFLSAFAVCDAADEP ALSALPEHVKPLANTLSSPHIQISVAPKKPSGILMG RFSQLLSFETEEPISQPSEEELAARLCTQEIVQKC QINNIFTESKFLQADSLMHLVRAIVGAEQIQAETG YPEDEANATFCLELLIAITLNNRDRISLLWQSVYE HISNIVQSTMAPCALVEKAVFGLLRICQRLLPYKE NLIDELLKSLQLVLKLDARVADAYCEQITQEVGQL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
566	sp_018890_pzwn	-0.369023	-0.927159	-0.365884	PELOTA (PEL1); FUNCTIONS IN: translation release factor activity; INVOLVED IN: meiosis, translational termination, translation; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: eRF1 domain 2 (InterPro:IPR005141), eRF1 domain 3 (InterPro:IPR005142), eRF1 domain 1 (InterPro:IPR005140), Probable translation factor pelota (InterPro:IPR004405); BEST Arabidopsis thaliana protein match is: Eukaryotic release factor 1 (eRF1) family protein (TAIR:AT3G58390.1); Has 776 Blast hits to 774 proteins in 323 species: Archae - 255; Bacteria - 0; Metazoa - 140; Fungi - 159; Plants - 51; Viruses - 1; Other Eukaryotes - 170 (source: NCBI BLink).	MKIVRRDYVHGGPGSIKMIPDEADDMWIAYNLIA EGDSVLAVTVRKVMREAASGGRDAERVKLRLEI VVEAVEYEKEGSVLRIRGKNTLENDHVKIGAFHT LELELHRPFVLRKDIWDNLALDMLYQAADPAASA DLVVILMQEGLAHLLLVGKSVTLTRSRIEVSIPRK HGPAIAGYESALKKFFENVLQAFLKYYDFNVVRC AVIASPGFTKDQFYRHLLLEAERKQLRSIIENKSRI VLVHTTSGYKHSLREVLDAPNVMSMIKDTKAARE VRALEDFFNMLSNDPARACYGPKHVEVAHERMA IQTLLITDEQFRNADIATRKKYVNLIGSIKASGGSV HIFSSMHVSGEQLAQLTGVAAILRFPLPLDDIEM
567	sp_019490_aoeg	#N/A	-0.837825	#N/A	CBF1-interacting co-repressor CIR, N-terminal;Pre-mRNA splicing factor; CONTAINS InterPro DOMAIN/s: CBF1-interacting co-repressor CIR, N-terminal (InterPro:IPR019339), Pre-mRNA splicing factor (InterPro:IPR022209); BEST Arabidopsis thaliana protein match is: CBF1-interacting co-repressor CIR, N-terminal;Pre-mRNA splicing factor (TAIR:AT2G44195.1); Has 59661 Blast hits to 33745 proteins in 1751 species: Archae - 130; Bacteria - 4372; Metazoa - 27826; Fungi - 6422; Plants - 3866; Viruses - 275; Other Eukaryotes - 16770 (source: NCBI BLink).	MALKFLNKKGWHTGSLRNIENVWKAEQKNEAE QRKLEELRKQIIEERERDEFRVLQEQAGLTPRQE RLEFLYDSGLAVGKPSSDGFQALGSFPSADNSA SASASAPVLASAPAPSSEAQPAVPGALFEEKPQ SANDAWRKLHSDPLLMIRQREQEAMARIKNNPV KMAMIRKSVESKKKKHDESHDGKHQKVKSKHKKH SSSSKHHSESESDSHEERERRRKVKQQRKEHH SESESDSHEERERRRKVKQKRKEHHSDTESDEE RERRRKTKNKKKEHSDSDEDDSRDRTRHHKHS HRNYDGKQKSDRNKDSDRIQRGYENSKHSEYES KHTGLPPRQDYHESKHTALPPRQESHRRRGRAPQLSEEE PRQDYHESKHTALPPRQESHRRRGRAPQLSEEE RAARLREMQMDAETHEEQRWKRLKKADENDAR EASQAGNSRGGKSFLDSVQKSVYGAEKGGSSTI EESVRRRSHFLQGRSSADEGNAFRR
568	sp_019500_ixjs	#N/A	-0.844228	-0.386128	phosphoprotein phosphatase inhibitors; FUNCTIONS IN: phosphoprotein phosphatase inhibitor activity; INVOLVED IN: regulation of signal transduction, regulation of phosphoprotein phosphatase activity; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein phosphatase inhibitor 2 (IPP-2) (InterPro:IPR007062); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MKGRVRWDETNLGDIEKNKPVRQKITEPKTPYH PMIEDDGSLSPIRSAFDECIDEAEHAKAIRHALND VAKDVASSSGKKSPRSGGWTSSEDERDVMEQD DEDSDLERSMSFKELRRAHYDEFLKVKELRRKG SFLEEEYDENGNHIKNINNNDCSSSLSPAVKAIDI HQGNPDETATHPPSTNGV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
569	sp_020010_coop	-0.8814	-0.884256	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 16 growth stages; Has 42 Blast hits to 42 proteins in 17 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 41; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	MSSTNAAKSWLQSLRRYIKAPWQITGPCASPEY KSAVPRATEYRLFCPATIPQEAIVPSSNPDTVFDI KYFPRDQRRNRPAIRRTVLKKADVEKMMAEKTF QPSDYPIPYLTATIEEDMNTHGGGYQ
570	sp_020070_xujw	#N/A	#N/A	-0.880398	SMAD/FHA domain-containing protein ; CONTAINS InterPro DOMAIN/s: SMAD/FHA domain (InterPro:IPR008984), Forkhead-associated (FHA) domain (InterPro:IPR000253); BEST Arabidopsis thaliana protein match is: SMAD/FHA domain-containing protein (TAIR:AT3G07260.1); Has 459 Blast hits to 458 proteins in 156 species: Archae - 0; Bacteria - 0; Metazoa - 122; Fungi - 202; Plants - 94; Viruses - 0; Other Eukaryotes - 41 (source: NCBI BLink).	MAHTAVAGGGGGSDVEAGFAKLQGEDFEYYMQ TYSIILGRNSKKSTVDVDLASLGGGMNISRHHARI FYDFQRRRFALEVLGKNGCLVEGVLHLPGNIPVK LDSQDLLQIGDKEFYFLLPVRNILGGAAVPRHHVY NTPLPPPAVPARNSGYGGGGGGSRKGRMKEFY AAPAAGEEYEEEEGGGRERMGEDGGGLSKKKMR RYGDGSGGGGHDGYGYGGGSGGKGGLSGTLE KKSDGRSRVDPEADNQQLMQLEEKDVVTSVATL LSDMCGPGDWMAMEKLHTELMEQYGHVWHHS RVRRYLTSEDWTGPESKGKPWYGLLMLLRKYPE HFVINTRSKGRVTLEFVSLVTLLT
571	sp_020340_amsg	-0.923033	#N/A	-0.883538	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSGREDDDSDAPEEVTALQGIQQDEEIRKALQLS KSRVRREGKERRKKLAARLTPRQKSKDEVPDGD VEMVEDDVEDDESKQSKVNTGMLPSDVVNLLAA REKQVFLSDSEDEKEEKKLKPKKKRAKFSGVETV ILKELPPPHCVQNSLEFLKKRKMQVPRSNAVLNN SKQALRLLSSSGLLK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
572	sp_020530_msda	-0.881977	-0.9601	-0.917146	protein kinase family protein / peptidoglycan-binding LysM domain-containing protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation, cell wall macromolecule catabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Peptidoglycan-binding Lysin subgroup (InterPro:IPR002482), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase-like domain (InterPro:IPR011009), Protein kinase, catalytic domain (InterPro:IPR000719), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: chitin elicitor receptor kinase 1 (TAIR:AT3G21630.1); Has 117626 Blast hits to 116415 proteins in 4493 species: Archae - 110; Bacteria - 13527; Metazoa - 43649; Fungi - 10139; Plants - 32288; Viruses - 471; Other Eukaryotes - 17442 (source: NCBI BLink).	MDHEGRGLDRVHSRTHKNSLCSLRIKSSNILLDGAFRAKISDLGLSKLVGTTSDEEASATRVVGTYGYLAPGS
573	sp_020760_mzey	-0.916596	-0.944215	-0.942535	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; Has 29 Blast hits to 29 proteins in 11 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 29; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MALRRALGWSDGELMRTDCKPCSRLMRQTAGIFTVGGALGFWVLCRLHYGTCNKCLI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
574	sp_020910_ppkh	-0.8894435	-0.8830224	#N/A	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G14750.1); Has 5478 Blast hits to 4354 proteins in 533 species: Archae - 87; Bacteria - 653; Metazoa - 2554; Fungi - 380; Plants - 418; Viruses - 16; Other Eukaryotes - 1370 (source: NCBI BLink).	MGSKGRLPPHIMRHPLHGPGLVHPEPFGAGIRP PHGGFPHGDMLPPPEILEQKLASQHVEMERLAT ENQRLAATHGSSRQQLAAAQQELQMLETQIRAS KSEREQQTRGLMDRISKMETELKAVDRLKLDLK QARTEAQTLVEVRQELITKAQQLNSELQRAHVDV QQIPAMMSELNHLRQEFHQYRTTYEHEKKVYRD HIESLQAMDKEYKSMADEVAKLRTELNNTSNVDK RSVYGSAAGYGDIDASAHNPSAQSSYEDGYGM QQVHHQFPASGAATTASTPAATGGGVVASSGG TPTYAPPQSGTASGRVGYGNPGYEAQRGLTYD PQGSQVYNNAQRASGYDGYNRGVNYDMQRPP YDPQRAAYEMPWANNYDAHSRGAAAAPPPPPS QGQAPVNNMAYGSAAPPPAARTTAGNEMQTRG GNTARR
575	sp_022320_njtt	-0.9111134	-0.9022964	-0.9228365	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT4G17616.1); Has 9904 Blast hits to 5133 proteins in 176 species: Archae - 0; Bacteria - 34; Metazoa - 111; Fungi - 31; Plants - 9502; Viruses - 0; Other Eukaryotes - 226 (source: NCBI BLink).	MVKMLRIKKLRCQILSSWLGGTRYVISSVLSNYR EINPSQPAISSFVNGSRLNPSLLYQGNVANRTKF FSTTAGTILVQARDPAKLSLDIHNAVNDHRYDDA WRYFEQHLEMEGFPRKSVVNKLLAGYAETLDSQ LLERAFGLVERAFEEGKHTLFEKVTLIYLAFALAR CSQPVPASVILRKLIRMEEYPPVTAWSAVLAYMS LSSSGAYLAAELILEIGYLFQDNRIDPRKKSNAPLI AMKPNTLAISVALAGCLVSGTTRIAEQLLDMMPRI GVKTDATLLILMAHIYERNGRREELKKLKPHIDEA HDLNDIQPRQYYNCLLNCHLKFGDLDSASRMVLE MLRKAKEANNSLAAATLRFEPANSYSPIKVSNITE LSISDKGGNQHFMSFEDFCRDRNFICLEGETKEV LISLSLKLQKQTELIMTERGILQPTEKTYVKLAKAF LEACKTKDLAEFLIKVEKEDSPFSVEDSPLVDVIN SCISLGWLDQAHDLLDEMCFAGIKISSSVYSSLLK AYCKEKRLSEVMSLLREVRKAGIQLDASCYDALI QSRVLQRDTQGALNLFKEMKETKIPRAGHQEYE VQARSCAKSSEAGFMAKLLQEIKEGETVDCGVH DWNNVIHFFCKKRLMQDAEKALKKMRSLGHTPN AQTFHSMVTGYAAIGGKYIEVTELWGEMKSLAAA TPMKFDQELLDSVLYTFVRGGFFFRANEVVEMM EKGNMFIDKYRYRTLFLKYHKTLYKGKTPKFETE AQSKKRDATLAFKKWVGLF
576	sp_022610_ktnh	#N/A	-0.9211137	-0.855111	uncharacterized protein	MADLYDPMVVPDTPDRGIAMTLSFSRKRKGNGD KVESSKSKGNGDGKDKVESSKRGKVSNDDVFE MLKGYYPSLERHMMGDSSDSE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
577	sp_022780_iqwn	-0.854134	-0.806465	-0.958837	dicer-like 1 (DCL1); CONTAINS InterPro DOMAIN/s: Restriction endonuclease, type I, R subunit/Type III, Res subunit (InterPro:IPR006935), Double-stranded RNA-binding (InterPro:IPR001159), Argonaute/Dicer protein, PAZ (InterPro:IPR003100), Ribonuclease III (InterPro:IPR000999), Double-stranded RNA-binding-like (InterPro:IPR014720), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021), Dicer double-stranded RNA-binding fold (InterPro:IPR005034); BEST Arabidopsis thaliana protein match is: dicer-like 3 (TAIR:AT3G43920.2); Has 21958 Blast hits to 17420 proteins in 2982 species: Archae - 328; Bacteria - 11461; Metazoa - 3615; Fungi - 1668; Plants - 1373; Viruses - 45; Other Eukaryotes - 3468 (source: NCBI BLink).	MEGENKVSDDSRPNYWLDACEECDDIPCEPIGD FGEFGSCNQPEVAGIDPSFFGGIDHILDSIKNGAG LPIAADNNDGMDIITPPPTTPKDMQVENLVVNCN VEVQRSSDNKLNGVDNGSNNMNIDNDNDNMMN MNNGDGDKGVVVDVRTSNGDSKRPSAREDCNV DRDRDSEERGGKRTRLSSGFRDERGYSSRVQS YGESNGRKRYREFDETGRRDRDRDRDRDRDLGGR RREYNNNGRRESRDRERRDREGKGYWEREGP GMKEIVYRVGSYEPARNRDANKVEFGKNQECNV RVGKNEEGNNKEKNLEERARHYQLEVLEQAKQK NTIAFLETGAGKTLIAVLLIKSICNRLQQQNKKVLA VFLVPKVPLVYQQAEVIREQTGYSVGHYCGEMG QDFWDTRRWQREFETKQVLVMTAQILLNILRHSI IKMEYIHLLILDECHHAVKKHPYSLVMSEFYHTTP KEKRPAIFGMTASPVNLKGVSSQVDCAIKIRNLET KLDSVVCTIKDRKELEKHVPMPSEIVVEYDKAASL LYLHEQITQMEVTVEEAARASTRKSKWQFMGAR DAGAKEELRQVYGVSERTESDGAANLIQKLRAV NYALGELGQWCAYKVAHNFLVALQNDEPAINYQL DVKFQESYLSKVVALLQCQLSEGAAFNKDASSTE EVNGFHQDNGSVDEIEEGELPDCHVVSGGEHVD VVIGAAVADGKVTPKVQSLVKVLMKYQHTDDFR AIIFVERVVAALVLPKVFAELPSLNFIKCASLIGHN NNQEMRTSQMQDTIARFRDGRVTLLVATSVAEE GLDIRQCNVVIRFDLAKTVLAYIQSRGRAPKPGS DYILMVERGNMSHETFLRNARNSEETLRKEAIER TDISHLKGTAGLVSADNVPGSVVYQVASTGATVSL NSAVGLIHFYCSQLPSDRYSILHPEFIMERHEKPG GSTEYSCRLHLPCNAPFENLEGPLCSSMRLAQQ AVCLAACKKLHEMGAFTDMLLPDKGSGEDGEKI EPNDEGEPLPGTARHREFYPEGIASILRGEWILS GRDGCDNSKIFHLYMYVVKWEHVGSSKDPFLTQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
578	sp_023010_ecqs	-0.917228	-0.909591	-0.922745	Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: transmembrane transport; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: SPX, N-terminal (InterPro:IPR004331), Major facilitator superfamily MFS-1 (InterPro:IPR011701), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein (TAIR:AT4G11810.1).	MVAFGKKLKERQLQEWQGYYINYKLMKKKVKQY GQHIETGTLGRRQVLKDFSRMLDIQIEKIVLFLLE QQGLLAGRLAKLGEQHNSPDQQSDIATISQLREA YREVGLDLLKLLYFVEINAIGLRKILKKFDKRFGYR FTDYYVKTRADHPYSQLQQVFKHVGIGAVVGAIT RNLGDLQERQGSYASIYDQPALPLQDPVIDSIKSA VDRLGYSTNFLHFLAQHALIMQDEILPSPVEEELV DKSYHFDSLILNLANTFLYMINTYIIVPTADDYSMS LGAAATVCGIVIGAMAVAQVFSSVYFSAWSNRSY FRPLVFSSIVLFVGNIMYALAFDFNSIWILLIGRVF CGLGSARAVNRRYISDCVPLKIRMQASAGFVSAS ALGMACGPALAGLLQTNFKLFNLTFNQDTLPGW VMAVAWLLYLVWLWIAFKEPSHEDDYVEQQSNT ETQQNDVMERGLQQPLLITEKIERSSEEKERNGD DDDGDAECDESEEAPEESRKPVTSFASAYRLLT PSIKVQLLIYFMLKYAMEILLSESSVVTTYYFGWT TSSVSIFLACLGLTVLPVNIVVGSYISNMFQERQIL LASQIMVCIGILLSFNFFVKYTVPQYVCSGLIMFVS AEVLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEA GTLARVVADGTITLIGLLGTSSLLNLTLLPSFLICV GCIIATCYTYNSLY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
579	sp_023490_topj	#N/A	#N/A	-0.848471	PKDM7D; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Transcription factor jumonji/aspartyl beta-hydroxylase (InterPro:IPR003347), FY-rich, C-terminal (InterPro:IPR003889), FY-rich, N-terminal (InterPro:IPR003888), Transcription factor jumonji, JmjN (InterPro:IPR003349), Zinc finger, C5HC2-type (InterPro:IPR004198), FY-rich, C-terminal subgroup (InterPro:IPR018516), Transcription factor jumonji (InterPro:IPR013129), FY-rich, N-terminal subgroup (InterPro:IPR018518); BEST Arabidopsis thaliana protein match is: Transcription factor jumonji (jmj) family protein / zinc finger (C5HC2 type) family protein (TAIR:AT1G30810.2).	VVVNAVFQQLGLVNLIFRYFRSRLSIMGTELLRAC LKEEDLDFLAVPPGFESFKAKKVQNADSEPCCSV SSNLSQQESVNVNSEPGIGGGGKISRSLRHRPS MNYGRYNQCSDEESDCEKLDQNFPTKPGLPKG VIRGCTECQNCQKVIARWHPEAACRADLEEVPV FYPSEEEFKDTIRYIEKIRPQAELYGICRIIPPPLWK PPCPLKEKDIWEHAKFVTRVQRVDKLQNRNASR KQSKVHSHSRKKRRRCGRMDVEFGSGNGKTSE PGEAEFGFEPGPEFTLDEFQRYADTFKVQYFRS ENCTDSGVENWEPSVDNIEGEYWRVVEKPTKEI EVLYGADLETGVFGSGFPKFSDQGCSTSDEDYV KSGWNLNNLPRLAGSVLSYENADISGVQVPWLY VGMCFSSFCWHVEDHHLYSLNYMHFGAPKLWY GVPGSDAIKLEEVMKKRLPDLFDEQPDLLHKLVT QLSPSILKSAGVPVYRCVQNAGEFVLTFPRAYHA GFNSGFNCAEAVNVAPVDWLPHGQNAVELYRE QGRKTTISHDKLLLGAAREAVKANWELSLLRKNT MDNLRWKDVCGRDGILAKSLRARVDTENVRREF LCKFSKALKMENSFDATTERECSICFFDLHLSAA GCRCSPDRFACLNHAKQLCSCAWDGKFFLFRYD VSELNLLADALEGKLSAIYRWARLDLGLALSSCIS KDNGQLTGLSGAVSEQPGENIICKTNSQPIVTLK NNEVIAKTPAVLASSVGKMSLVEKGKFPDAVLAL EATKVAPIKLSRQSKRKSDKQQIGKDGEEVPNKR FTGMEPQLKLSSHGPEDLICLSDDEDEQSNKTYI DANKGTIGQDSPNVPEIIEGSDHKLNLCDLANAAL AAEANLNLISDGHKDDCMSCSTSPKEETNVKCET VAVSDEHNLSCHQGSIGSSIEKKDPDASTPDSFL QNAACVPNVGCGIKFRDESVVQRVLDDARILPGS PSCSQSSVDKYPRQKGPRIAKVVRRINCNVELLE FGVVIPGNLWSTSRAIFPKGFKSRVRYINISDPTN MCNYISRILDGGLNGPLFMVFLEHCPSEVFVQTL
580	sp_024110_wuro	-0.909107	-0.862061	-0.889132	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: TATA-binding protein binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MAEIEELEAQNPEIDSIKAAKKAVKDEERERKRRE KRRKILMKESEKADKRGVCYLSRIPPKMDPLKLR QTLSPYGEILRIYLAPEDPEAQTRRKRSGGFRGQ EFKEGWVEFADKRIAKRVANMLNGEEIGGKKRS KFYYDIWNIKYLSKFKWDDLTEETTIKNATREQRL RMELSSANKEQDFYLSKVDQSRAFSEIEKRMAK KRKLENGVDSEGPNEQQKPKVIRPFSQKKPIAEK AEKDKPKLSKAVLAEIFNSGNGEA

PCT/IL2019/051000

	A	B	C	D	E	F
1		**correlation coefficient with baits used in the analysis**				
2	**coexpressed gene**	**sp_107620_kqnh (*SOAP1*)**	**sp_107660_kiqg (*SOAP2*)**	**sp_107670_ptqx (*SOAP2-like*)**	**annotation**	**aa seq**
581	sp_024490_etgk	#N/A	-0.966627	#N/A	Transducin/WD40 repeat-like superfamily protein; CONTAINS InterPro DOMAIN/s: WD40 repeat-like-containing domain (InterPro:IPR011046), WD40 repeat (InterPro:IPR001680), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); Has 672 Blast hits to 646 proteins in 166 species: Archae - 0; Bacteria - 97; Metazoa - 271; Fungi - 156; Plants - 85; Viruses - 0; Other Eukaryotes - 63 (source: NCBI BLink).	MGVFAVDTFDLADVEWSLDDSSIVVWDSLLEYK VLIYTPDGRC
582	sp_024650_kuuy	-0.933326	#N/A	-0.863408	uncharacterized protein	MQFSKAGLDVSVSDMGTKRKKVGSCESEVDAS QCSPFEVAAFKPVLIETVATNGFTYSVPSHGPDE QHSVDDLGGTEPEELIEVAIGDSCKVMNLDDGNS SHIQCPDTDKKSVAAEVSISAGLYGCLLISLILVQC HRISLHLRCHASYRRVFKNLCCKGKVD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
583	sp_024780_ydyr	#N/A	#N/A	0.87921	calcium-dependent protein kinase 1 (CDPK1); FUNCTIONS IN: calmodulin-dependent protein kinase activity, protein kinase activity, kinase activity; INVOLVED IN: response to water deprivation, response to salt stress, protein amino acid phosphorylation, N-terminal protein myristoylation, abscisic acid mediated signaling pathway; LOCATED IN: chloroplast; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), EF-Hand 1, calcium-binding site (InterPro:IPR018247), Serine/threonine-protein kinase domain (InterPro:IPR002290), EF-hand-like domain (InterPro:IPR011992), Calcium-binding EF-hand (InterPro:IPR002048), EF-hand (InterPro:IPR018248), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), EF-HAND 2 (InterPro:IPR018249), Calcium-dependent protein kinase (InterPro:IPR020642), Calcium/calmodulin-dependent protein kinase-like (InterPro:IPR020636); BEST Arabidopsis thaliana protein match is: calcium-dependent protein kinase 30 (TAIR:AT1G74740.1); Has 125370 Blast hits to 119826 proteins in 3847 species: Archae - 176; Bacteria - 14154; Metazoa - 46310; Fungi - 16874; Plants - 24900; Viruses - 438; Other Eukaryotes - 22518 (source: NCBI BLink).	MGNCNACIRPEELPKKTPQTKKKKTRPNPYSINQ PQPAPVRVLKDIIPAVGPRISDKYVIGRELGRGEF GITYLCTDRETKDRLAVKSISKRKLRTAIDVEDVR REVAIMSTLPVHPNIVRLKAVYEDNEAVHLVMEL CEGGELFDRIVARGHYSERAAATVAKTVAEVVK MCHDNGVMHRDLKPENFLFANKKENSPLKAIDF GLSVFFRPGERFTEIVGSPYYMAPEVLRRNYGQ EVDVWSAGVILYILLCGVPPFWAETEQGVALAIL RGVLDFKREPWPQISDSAKSLVKQMLDPDPKKR LTAQQVLEHPWLQNAKKASNVPLGDIVRTRLKQ FAGMNRFKKRALRVIAEHLTVEEVEIIRDMFILMD TDRDGKVTYEELRTGLKKVGSQLGEPEIKLLMEV ADADGNGYLDYGEFVAVTIHLQKMENDEHFHRA FMYFDKDGSGYIDLDELRDALSDESAETDADVLT EIMREVDTDKDGRISYEEFVAMMKAGTDWRKAS RQYSRERFKSLSLNLMKDGSLQLHDGMTGEAVV V

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
584	sp_024870_skya	#N/A	#N/A	-0.8731188	BRI1 suppressor 1 (BSU1)-like 3 (BSL3); FUNCTIONS IN: hydrolase activity, marganese ion binding, protein serine/threonine phosphatase activity, phosphoprotein phosphatase activity, iron ion binding; INVOLVED IN: biological_process unknown; LOCATED IN: cytosol, plasma membrane; EXPRESSED IN: guard cell, cultured cell; CONTAINS InterPro DOMAIN/s: Serine/threonine protein phosphatase, BSU1 (InterPro:IPR012391), Metallophosphoesterase (InterPro:IPR004843), Galactose oxidase/kelch, beta-propeller (InterPro:IPR011043), Kelch repeat type 2 (InterPro:IPR011498), Kelch-type beta propeller (InterPro:IPR015915), Serine/threonine-specific protein phosphatase/bis(5-nucleosyl)-tetraphosphatase (InterPro:IPR006186); BEST Arabidopsis thaliana protein match is: BRI1 suppressor 1 (BSU1)-like 2 (TAIR:AT1G08420.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MEVDSSSSMENDQDPSSTQNNNSSSPSSPPPL PSSSSPLSSVQTTPASSSPMEKDVEQQQLRQQ CQGQIVGPRCAPSYSVVNAIIEKKEDGPGPRCGH TLTAVPAVGEEGSTGYIGPRLILFGGATALEGNS AAAGTPSSAGSAGIRLAGATADIHCYDVLTNKWS RIAPIGEPPTPRAAHVATAVGTMVVIQGGIGPAGL SAEDLHVLDLTQQRPRWHRVVVQQPGPGPRYG HVMALVGQRYLMCIGGNDGKRPLADVWALDTAA KPYEWRKLEPEGEGPPPCMYATASARSDGLLLL CGGRDANSVPLASAYGLAKHRDGRWEWAIAPG VSPSPRYQHAAVFVNARLHVSGGALGGGRMVE DSSSVAVLDTAAGVWCDTKSVVTSPRTGRYSAD AAGGDASVELTRRCRIHAAAAVGDLIFIYGGLRGG VLLDDLLVAEDLAAAETTTAASIAAAAAAAASNSPP GRLPGRYGFTDDRLRESDPESVPDGAVVLGNPV APPVNGDIYTDINHENAMAQGSRRLSKGVEYLVE ASAAEAEAISATLAAAKARQVNGEVELPDRGDR GAEATPSGKPIPTMIKPDSSAPISTPPPGVRLHHR AVVVAAETGGALGGMVRSLSIDQFENEGRRVSY GTPENATAARKLLDRQMSINSVPKKVVAHLLKPR GWKPPVRRQFFLDCNDIADLCDSAERIFASEPSV LQLRAPIKIFGDLHGQFGDLMRLFDEYGSPSTAG DIAYIDYLFLGDYVDRGQHSLETITLLLSLKVEYPH NVHLIRGNHEAADINALFGFRIECIERMGERDGIW AWHRFNRLFNWLPLAALIEKKIICMHGGIGRSINH VEQIENIQRPITMEAGSIVLMDLLWSDPTENDSVE GLRPNARGPGLVTFGPDRVMEFCNNNDLQLIVR AHECVMDGFERFAQGHLITLFSATNYCGTANNA GAILVLGRDLVVVPKLIHPLPPAISSPETSPDRHIE DTWMQELNANRRPPTPTRGRPQVANDRGSLAWI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
585	sp_025100_pcos	-0.8841178	#N/A	#N/A	Nucleotidyltransferase family protein; FUNCTIONS IN: nucleic acid binding, nucleotidyltransferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: 4 anthesis, F mature embryo stage, petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: Nucleotidyl transferase domain (InterPro:IPR002934), PAP/25A-associated (InterPro:IPR002058); BEST Arabidopsis thaliana protein match is: Nucleotidyltransferase family protein (TAIR:AT4G00060.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MAVTKYSYETLSPLLPSSAADTTATIAGNDEYDD FVVFRNEISLSSLEPTSTDAAAVDYFSLDVNDFSD KVIQTNPSPEVVVAAERDKEAEQIPERTLESGWF RADCRFKSPMLQLHKEILDFCDFLSPTPEEAASR NSAIERVSDVIRYIWPYCRVEVFGSYRTGLYLPSS DIDVVILESNIKTPQIGLQALARALSQKGIAKKMQV IGKARVPIIKFVEKVTDVQFDISFDVDNGPKAAEYI QDAISRLPPLRPLCLILKVFLQQRELNEVYSGGVG SYALLAMLIAMLRSVNDCQRSPEQNLGILLVKFFE FYAHRLNTWDVGVSCNGRGTFYLKSRKGFQSK GRPFLISIEDPQTPGNDIGKNSFNYFQIRSAFMMA YATLTNTKTIMNLGPNRSILGTIIRPDPVLLERKGG SKGELTFNSLLPGAGEALDSLYDEKQDILCNWRV DNDYEEEEEEPLPRGGNAESSGASSSGKKRKASK SKSYGKVKENGSFGKNNYEENGSRKERKRDKK RSKSSG
586	sp_025110_iqaf	-0.8855534	-0.1423932	-0.8933853	uncharacterized protein	MFAFRRRQSQQLNRRILSPPVVDLRSPLTTRHSS LATRRAARRSQQSGDFTVIWCDWFRLLAWSGC
587	sp_025140_mazi	-0.8853776	#N/A	-0.8844138	ATCOAE; FUNCTIONS IN: ATP binding, dephospho-CoA kinase activity; INVOLVED IN: coenzyme A biosynthetic process; LOCATED IN: peroxisome, chloroplast, vacuole; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Dephospho-CoA kinase (InterPro:IPR001977); Has 7074 Blast hits to 7073 proteins in 2608 species: Archae - 1; Bacteria - 4684; Metazoa - 251; Fungi - 135; Plants - 58; Viruses - 0; Other Eukaryotes - 1945 (source: NCBI BLink).	MRIVGLTGGIASGKSTVSKLFKAQGIPIVDADIVAR DALKKGTGGWKKVVDAFGDGILQPDGEVDRPNL GRIVFSNPDKRQLLNRLLAPYISSGIFQEILKLWLK GHKVIVLDVPLLFEAKMDKWTNPVIVVWVDRET QLNRLVSRDRITEEEANNKIDAQMPLDSKKDKADI VIDNTGSLDDLNEHLQNVLRVVNRPLTWSEFVLS RQGAFSAFAFVVVGVFFVAKNIFKCRL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
588	sp_025650_iaes	-0.885419	-0.824133	-0.953082	Homeodomain-like transcriptional regulator; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent, regulation of transcription; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: DDT domain superfamily (InterPro:IPR018501), Homeobox (InterPro:IPR001356), Homeodomain-like (InterPro:IPR009057), DDT domain, subgroup (InterPro:IPR018500), DDT domain (InterPro:IPR004022), Homeodomain-related (InterPro:IPR012287); BEST Arabidopsis thaliana protein match is: homeobox-1 (TAIR:AT1G28420.1); Has 67503 Blast hits to 37811 proteins in 2117 species: Archae - 251; Bacteria - 7621; Metazoa - 32181; Fungi - 4940; Plants - 2680; Viruses - 355; Other Eukaryotes - 19475 (source: NCBI BLink).	MEGGSCSDGERKPPESGGGGGGGGSGGGEPKT KRKMKTASQLEVLEKTYQVATYPAEALRAELSVQ LGLTDRQLQMWFCHRRLKDRKPQQGGGGGGGG GGGGGGSGKRSKKREASPAAAAATGGPNVAVAP LLPPAPAPVGLRDEMLGAELGNEHGSGSGSGSS PYGHMDPRRGYVGRPGVAVPRISGHGRYYEPS QSVAELRAIAFVEAQLGEPLRDDGPALGMEFDPL PPGAFGAPITTTTAPQKPTGRTYEYDRADAKPVK GSNRAPHEYQFLPEQPTVRSESYERVAQSHYYG SPSDGPGTKTPPLSSGRSYIHGYEEVPVVYGVP NQVSSLNLMSQGKVGSLMPSTSGDYETIARKSP YMNMGMESPFSAHPTSGILDNALISSERRATQEE DAGRLERKRKNEEARVAKEIEAHEKKMRKELER QDILRKKREEQLRKDMERQDRERKKEEERLLRE KQREEERYQREQKREMERREKFLLKETIRVEKM KQKEELRREREAARQKAANDRAVARRLAKESTE LIEDERLELMEIAASRKSLTSILSLDSETLQNLDSF REMLRVFPPKSVQLKKKPFATRPWMGSEENVGN LLMVWRFLVTFADVLGLWPFTLDEFVQAFHDYD PRLLGEIHVSLLLRTIIKDIEDVARTPSTGLGVNQNS VANPVGGHPQIVEGAYAWGFDIRSWQCNLNPLT WPEVLRQLALSAGFGPKLKKRNTEQAYSREEYE GNDGIDIISNLRTGAAVESAVAIMRERGYSNPRRS RHRLTPGTVKFAAFHVLSLEGGNGLNILEVADRI QKSGLRDLTTSKTPEASIAAALSRDTKLFERTAPS TYCVRSPFRKDPADGEAILSSARREKIREFKIGFVD GVEADVDGMEADDVEKDEDSESDIAEDPEIDDL GTELTPNNDVNPLYDGNSSVPKVESDNGSRDLH SHVVDKTPNGLADAGNVLSSAHLESFKELNDNG CFAEESMDVGGICQETNSTDHEDTDIDESCPGE PWVQGLMEGEYSNLSVEERLNSLVALLGVATEG NTIRVVLEERLEAATALKKQMWAEAQLDKRRLKE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
589	sp_025820_mrts	-0.864925	-0.951793	-0.921175	CONTAINS InterPro DOMAIN/s: Symplekin tight junction protein C-terminal (InterPro:IPR022075); BEST Arabidopsis thaliana protein match is: HEAT repeat-containing protein (TAIR:AT5G01400.1); Has 593 Blast hits to 297 proteins in 111 species: Archae - 0; Bacteria - 3; Metazoa - 139; Fungi - 49; Plants - 57; Viruses - 0; Other Eukaryotes - 345 (source: NCBI BLink).	MEETGLRVPEHSYMLVPQLLALLKDEDPGVARQ AIVSGTKSFCSILQELAQQFHRHGKVERWAEEM WMHMIEFKDAVLSLVFEPAFVGTRLLALKFLETCI LLLTSDASETERLTSGGKELRFNVSWLASGHPVL DAAKLISEANRLFTSLLDMLQSSSSLPGQLTITIVN CLASIARKRPEHYGQVLPLLLDFHSNSEIRGGHA ASIQYSVRSALLGFLRCTHPVIIESRDRLVQELRL MSAGDAVDQVLRQVDKQMRNNERSSRDARVGK EEHSSTQRPLSGDSIKRKLLPVDNEEPDNVFEMA SKRMRYTSNGKSPDLVQIHGSGDDNPVVNGISS EAPVVDNSVNPIEQMIAVIAALLAEGERASESLEL LISQIHPDLLADIVITNMKHLPKTAPLAKQGILPSN SQSSSSVSAAQGADIMSLPVHSSLSAHSAIPAAV AVSSSLPDFSGSANLQVDARRDPRRDPRRLDPR RVAVPAEATLLPNAEEDTHVSSVFEGPASVNRST SLTSALNTEPYSVPDAEVDGKSIRTLLVPEKEIMP KEEEADESKEFLSSSEFNVLVNGAHSPDPTVNES SITLTSLKMEPTDAASALTFQESEDDSPPASNTS ACEDTCEDLPEVPSYVYLTEKQRQIVKKLAVERIV KSYKQLKDQEFSKTRMSLLARLVSWMGGDDLVT MMVQQDIVLDYRDQQGHDLVMHVLYHLHGLITIE VEQSSGAADLYEKFLLGVAKSLLDTLPASDKSFS RLLGEAPMLTDSVTRLLENLCYPDGSDMRGKDIR DDRVTQGLGAVWSLILGRPLNRQACLSIVLKCAV HSEDDTRAKAIRLVSNKLYQLNYVSETIEQYAKN MLLSAVDCTTDADASESICSGREVGVAEETSVTG PQLIDFGTSHAVSQAVASMSFPEAQRLISLYFSLC TKKPGLLQIVFDVYARAGKIVKQAIHRHIPILLRAL GSSYSQLLRIISDPPMGSENLLMLVLQILSEQTTP PPDLISTVKHLYETKLKDATILIPMLSSLSKNEVLPI FPRLVDLPLDKFQQALAHILQGSAHTGPALTPAE VLVAIHDIVPEKEKIALKKVTDACSACFEQRTVFT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
590	sp_026180_qgyj	#N/A	#N/A	0.850284	FAR1-related sequence 5 (FRS5); FUNCTIONS IN: zinc ion binding; INVOLVED IN: response to red or far red light; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: MULE transposase, conserved domain (InterPro:IPR018289), Transcription factor, FAR1-related (InterPro:IPR004330), Zinc finger, PMZ-type (InterPro:IPR006564), Zinc finger, SWIM-type (InterPro:IPR007527); BEST Arabidopsis thaliana protein match is: FAR1-related sequence 3 (TAIR:AT2G27110.2); Has 1793 Blast hits to 1580 proteins in 47 species: Archae - 2; Bacteria - 0; Metazoa - 4; Fungi - 136; Plants - 1646; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MENLIRSVIQSTQTIQVPVEPDIQDATMEDVLYNC IENNALAQEREEEEAMEAEEQDIISDSEEPDQDIV GTQMGYTAETVEELLGFYQKHASAVGFSIRKGN TRFKVGTRIVLEKTYVCSAAGVTNNGKNKKTKVQ TVVPVVPKKERKPRQVSITRTQCRACLRVKMNS EGRYEVVNHVIMHNHDLTRSQWHYLHRSERQIT EEKREAIETMQKSGNSYILIVKLLCSV
591	sp_026370_zdwu	#N/A	-0.874001	#N/A	unknown protein; Has 19 Blast hits to 19 proteins in 8 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 19; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAIISPVKEVSLELFRELLIAISQSEPENRVSSKSF YEYCNSLNGAKVKNEDKTEELRSELISISCASSAD NVIMPSMSMQEAES

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
592	sp_026920_oruh	#N/A	#N/A	-0.6773345	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Nop14-like protein (InterPro:IPR007276); Has 69842 Blast hits to 35213 proteins in 1572 species: Archae - 363; Bacteria - 20593; Metazoa - 20851; Fungi - 8010; Plants - 2912; Viruses - 517; Other Eukaryotes - 16596 (source: NCBI BLink).	MAKSETMKKNKKKSKKAGVKKVSMKQKSAPKD NPFETIWSRRKFDVIGKKRKGEERRIGQARTRAI DKRKKTLLVEYEQSGKASVFVDNRIGEQNDALG EFDKAILRSQRERQMKIKKTSKYNLSDEEEDDLDI YDGLRDDYDDEVEHGSDEDAGQDSSSILRHGRK SVDANIAEEENRLKSKKEVMEEIIQKSKFFKEQKA REKEENEKLVEQLDKDFESLVHSNALSSLTQPNK MNALKALVNPNFSKELKGKEVMSVSKVSKQDEP DTYEKLVSEMAFEMRARPSDRTKTAEEIAEEEKE RLERLEQERQDRMRATDDASEEDDDSESDGKP FTNKHSHVSGDDLGDSFTVDDKSLNKKGWVDEI LERNDTEDTEDEDEDSRNSDDSEESEDDEGGTE DEEDDGKAETLKDWELSDDEVIGTDLLDEEDKS DEEGGEEDNKTHSEIEDKKNERLANPVRKVKKK TALLDLGMAQRSSLKQELPFKINAPGTMEELHAL LDDHSNEDIIEAIRRIRICNAIALAAENRKKMQVFY GLLLQYFATIPNKVPLNFQLLNLLVKPLMEMSSEI PFFSAICARERIQRTRTKFSGDIKDPGNSCWPSM KTLLLLRLWSMIFPCSDFRHVVMTPASLLMCEYL MRCPVLSVRDAAIGSFLCSIVLLVTKNSKKFCPEA LTFLKDLLLAATVPRCPKDSQMRTQDHKPLLWIG ESIEDIKPLDFLKIMNLPKDSSFFNSHDFRASVLVS VVETLRGFVNIYGELKAFPEIFSPIKALMIEVADQK NMPNPLQEKIRDVVELIKKKADEYDEHRQPLEMR KQKPVPIKLLNPKFEENFIKGRDYDPDRIRAENRK MKKLLRKEARGAVSELRKDNHFLQAVKEKEKQR VDEERAETYGKARAFLQEQEHNFKSGQLGKGRK RRR
593	sp_026950_rcux	-0.8871173	#N/A	#N/A	uncharacterized protein	MVESTSNSRTRDLDRLLLRSGHLVGPNFEPSAE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
594	sp_027370_yznu	#N/A	#N/A	-0.848904	HIT zinc finger ;PAPA-1-like conserved region; CONTAINS InterPro DOMAIN/s: Zinc finger, HIT-type (InterPro:IPR007529), PAPA-1-like conserved region (InterPro:IPR006880); BEST Arabidopsis thaliana protein match is: PAPA-1-like family protein / zinc finger (HIT type) family protein (TAIR:AT3G06660.1); Has 3160 Blast hits to 2461 proteins in 330 species: Archae - 7; Bacteria - 184; Metazoa - 1235; Fungi - 390; Plants - 202; Viruses - 26; Other Eukaryotes - 1116 (source: NCBI BLink).	MEDLNSSRFEGIGNAVKKRRSHRRPRPESQPLS DGRDVSSLSSTPPSDDVKPGSSDENSGGGDHS SKRKELSLSQCASRFPGPSGVEGENPVRKSKKV DIGLDELYGDSSNHKRSSEGVLAPGNWRSTSRE KESPEPQKTMALNKKADEGTGPSQDRLGNDNK VKKVKLKVGGVTRTIDANSKSSRSLDATIVLQKSA VKDNSEIDNSPSGDKKGGLQGIPWKDFSKVGFA LVREESSMKGGGKNSSVKQGDKSDSARKSKRS SKKRAIDGFDDDDDDEIRYLEKVKNAKYFTGFGE DDEESSRKHRRLSKFSGNTPQNSEVGASSSRLG KDGKKRSPSENTEFEGEEELGSDGELEGKKKKK ARKEPVDVIVEPKRELTLTTRQKALQSGRDASAP GGSAVEFPDGLPPAPSRKQKEKLSEVDQQLKKV EAAQRRKMHNEKIARENEAEAIRRILGQDSNRKK REDKKKKRVEELAQEREADELLRASNSIRLVMGP NGTTVTFPNEMLPSFFDSKPCSYPPPRESCAGP SCTNPYKYRDSKTKLPLCSLQCYKAVQGNVPTQ AATC
595	sp_027940_ssgd	-0.866782	-0.944894	-0.86951	ARGONAUTE 4 (AGO4); CONTAINS InterPro DOMAIN/s: Domain of unknown function DUF1785 (InterPro:IPR014811), Stem cell self-renewal protein Piwi (InterPro:IPR003165), Argonaute/Dicer protein, PAZ (InterPro:IPR003100), Polynucleotidyl transferase, ribonuclease H fold (InterPro:IPR012337); BEST Arabidopsis thaliana protein match is: Argonaute family protein (TAIR:AT5G21150.1).	MDKDDNNRVRSPMSRPGFGMKGRKISLCTNHF RVELTNSSGYFFHYSVSICNKDGHPVEGKGIGRR IMDKIHETYKLELAEKDFSYDGESSLFTWGALPS KAMDFTVVLDKISSKRNGDNSISNGSRSPTQGDK KRQKCLFQTRTFRVAIKCVGKIPAGAISDMLRGK YSEEVHEVIKVLDIVLRQRAAKQDCLIVRQSFFHD DQRNFTDIGGGVFACKGLHSSFCPTQNGLTLNM DSSRTTILKPGPVLDFLVSNQNVKNPSCIDWSKA RRMLKNLRIKIFPYDRESRITGLSEYVCKQQRFPL GDEGGRGDESQKERVTVYDYFVKHKKIHLRYSG DFPCINIGKTKSKLYPIELCSLLPLQRYTKELSIPQ RLKFIESIRQKPLDKMTMLTEAWKTNNYAAEPML QFCGVSISSSFTQVEGRVLPAPKLIVGRGEEIFPR AGCWNLNRKMLAEPMSLHSWAVVSFRSRFDASI VCKDLAKFGEMNGIFMEPPICIIEENGRFRQAPPA IRVERMFEQLDMQLKPLRARLSKNKTWPDFIICLL PNKDSEIYGPWKRKTLIEFGISNQCISSSKLDMQY FNSLLLKINAKLGGLHSRLAVEWNCRVLSKKPTM IFGMDVWHGYHGVPSIASVVGSRCWPSLARYRA SVRSQSPKTEMIDSLFKPPPGNAEKDEGMIRELL VDFYSSSGKVKPSQIIIFRDGVSESEFKQVLNVEF HQILEACKRLDSDWNPKFTIIIAQKRHHTKFFQSG QPHQNVPPGTVVDSKVCHPHYYDFYMCPHAAFS GTARPIHYHILLDEIGIPSDELQELVHSLSYVYQKC TKAISIVTPIRYAHLAASKTSHFLQYATSEIFSGHT EAQVSDLPKLHKDVCKSMFFC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
596	sp_027960_taqj	#N/A	-0.8838131	-0.6933036	Galactose mutarotase-like superfamily protein; FUNCTIONS IN: isomerase activity, carbohydrate binding, aldose 1-epimerase activity, catalytic activity; INVOLVED IN: galactose metabolic process, carbohydrate metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Glycoside hydrolase-type carbohydrate-binding (InterPro:IPR011013), Aldose 1-epimerase (InterPro:IPR008183), Glycoside hydrolase-type carbohydrate-binding, subgroup (InterPro:IPR014718); BEST Arabidopsis thaliana protein match is: Galactose mutarotase-like superfamily protein (TAIR:AT3G61610.1); Has 1923 Blast hits to 1920 proteins in 763 species: Archae - 0; Bacteria - 1269; Metazoa - 39; Fungi - 135; Plants - 272; Viruses - 0; Other Eukaryotes - 208 (source: NCBI BLink).	SSTLKALMALTRLFFAGFVASLPRFLDSLPRMMM LLFTSFKLMYLYGAHVTSWKNEHKQELLFVSNKA IFKPPKPIRGGISICFPQFDTLGYVGAHGYARNRV WSVDPNPPPLPMTSPTNSYVDLLLKPSEEDLKIW PHSYEFFRLRISLGPGGELLLTSRVRNIDSKPFSFT FAYHTYFSVSDVSEVRVEGLETLDYLDNTKNRER FTEQGDAITFEGEVDKVYLSTPSKIAILDHERKRT YVIRKDGLLLDAVVWNPWDKKAKALADFGDNEYK HMLCVEAAAIERPITLKPCEEWTGRQEISTVPSS YCSGQLDPQKVLMGN
597	sp_028180_arxd	-0.9333476	-0.8856877	-0.9233612	3'-5' exonuclease domain-containing protein / K homology domain-containing protein / KH domain-containing protein; FUNCTIONS IN: 3'-5' exonuclease activity, RNA binding, nucleic acid binding; INVOLVED IN: nucleobase, nucleoside, nucleotide and nucleic acid metabolic process; LOCATED IN: intracellular; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: K Homology (InterPro:IPR004087), K Homology, type 1, subgroup (InterPro:IPR018111), K Homology, type 1 (InterPro:IPR004088), Polynucleotidyl transferase, ribonuclease H fold (InterPro:IPR012337), 3'-5'-exonuclease (InterPro:IPR002562); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G25920.1); Has 1067 Blast hits to 1067 proteins in 347 species: Archae - 0; Bacteria - 413; Metazoa - 196; Fungi - 59; Plants - 125; Viruses - 0; Other Eukaryotes - 274 (source: NCBI BLink).	MADDPSSHHAHVPLPSNLGGLVSDDAPLVPIHIV TEASQLPVEFLEPCPETKLVIGFDCEGVDLCRYG TLCIMQLAFPDAIYLVDAIEGGDMLIEACKPALES RYITKVIHDCKRDSEALYFQFGIKLQNVMDTQIAY SLIEEQEGRPRVPDDYISFVALLADPRYCGISYLE KEEVRVLLRQDPKFWKYRPLSELMIRAAADDVRF LLYIYHKMMEKLNAHSLWNLALRGELYCRCFCIS DDNLTEWPALPPIPDDLPVDTQAPEEEILSVLDVP PGKMGRVIGRKGASILSIKESCNAEILIGGSKGPP DKVFVIGRVVKQVRKAEAMIRGRMLDM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
598	sp_028670_sawg	-0.8611885	#N/A	-0.9837435	ATBET12; CONTAINS InterPro DOMAIN/s: Target SNARE coiled-coil domain (InterPro:IPR000727); BEST Arabidopsis thaliana protein match is: BET1P/SFT1P-like protein 14A (TAIR:AT3G58170.1); Has 338 Blast hits to 337 proteins in 95 species: Archae - 0; Bacteria - 0; Metazoa - 144; Fungi - 35; Plants - 128; Viruses - 0; Other Eukaryotes - 31 (source: NCBI BLink).	MSYRRETRASKIGLFDGYNSLEEGGLRASTSYSS DISEHENDKALDTLKDRVFFLKQLSGDIQDEVLNH NRVLDRMGNEMDASRGILSGTRDRFKMVFERKS GRAMCKLVGFLVAFFLLLYYIIRTLLYQSEETAKW LAIEAEEGSLDIPCEYEVETL
599	sp_028720_fiuc	-0.8262233	-0.9844039	-0.9436312	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF642 (InterPro:IPR006946); Galactose-binding domain-like (InterPro:IPR008979); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF642 (TAIR:AT2G41800.1); Has 306 Blast hits to 298 proteins in 21 species: Archae - 0; Bacteria - 4; Metazoa - 0; Fungi - 0; Plants - 302; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MASFNYLCVVILLMMINSRSYVAAQLAPVSVISTD GLLPNGNFEQGPNPSDLKKITIVGKYSLPKWVKK GLVRYVSGGHKRVGFTFNVTRGVHAVSLGQDTF IRVTYGSNTADFAFPSKPLYLSDKADTFAWGFQT AASTAYVLFSNPTGHPKYPACGPLIQAVAIKELPP PPTTTIDNNLIQNGNFEIGPYTLENFTTGVYIHPYT LNPVSPLPGWIIESRYTIRYINSKDFFIPSGSSAVE LLVGGKENAIAQVIGTTPKMYYNLTYTIGDANNG CIGSMEVQASADKETIKVKYESQGKGGVTHASF RFQAVSLSFTTRISFWSVFYHNTVDDPNNRFCGPV LDNVRVVLDP
					RNA-binding KH domain-containing protein; FUNCTIONS IN: RNA binding, nucleic acid binding; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: K Homology, type 1, subgroup (InterPro:IPR018111), K Homology (InterPro:IPR004087), K Homology, type 1 (InterPro:IPR004088); BEST Arabidopsis thaliana protein match is: RNA-binding KH domain-containing protein (TAIR:AT5G15270.1); Has 6723 Blast hits to 3876 proteins in 329 species: Archae - 3; Bacteria - 266; Metazoa - 4280; Fungi - 575; Plants - 1237; Viruses - 20; Other Eukaryotes - 342 (source: NCBI BLink).	MSLSLTPAKRQHSRSHENSNGKKFQKSASLNST KQPRKTSSGGGVFRVLCPSSRIESVIGKENNILS QIHQETGAKVRIDDSIPGCDEKLIFIMGLDKEAES SGQHNKNDNEDTKSSKEQGSSEDHNVNDADKE TPTNAEDLLSKKGNSSVQNALLLVFQNFYEGERE TEEPNEESNESAAVTFRLLVLADQVGCLLGKGG SIIKQMSAESGAQIRVLPRDKLPQCASSTDEVVQI TGLADAVKKAIRTVSNQILENALHDNDSFPSNTP GRTSQSHSNHVPPPESYFLHSLPSHGPSGGGFR DSADFHSGVPSLFPKFHENFMPGRMDVPDIVTF RLLCPDERVGGVIGKGGAIIRTLQHETGCEIKVLE GPSDGEERMIVSGPAHPDDRISAVQDAVLRVQN RVAKAVPDSKEKGVMAKLLVPSNQIGCLLGKGG SIIGEMRKVTGAYIRILGKDQMPKSASENEEVVQI NGEPEVVQDALLQITTRLQHHFFRDAFPSVNHPS NLAFSDRAPPFPSFMGREEVSPRGMFPKLGPSF RNTDYIGLPPLHGGFHPHDEHPPFLHDIHRHGAP NLSERSWHPQNLIEGGPMGLLDHGGPPHRRMS GFGGGSQLAVITNTTVDVVPSSIVPIIYGEDGGC LRQIRQISDAKITITDPKQGAPETVIIISGTPEQTHA AQSLIQAFVMLENESG
600	sp_028900_izxk	-0.8652355	#N/A	-0.9223857		

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
601	sp_029530_dujr	#N/A	-0.878457	#N/A	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Defects-in-morphology protein 1-like, mitochondrial (InterPro:IPR019190); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G57110.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MFQGATEDGKTMVYKYGPWLGASPHKRSSRLS FSEREKEKKMLDRLNTTHAPRVARHGGFNDPNA IRLEPPGVARKLSFSSPRSSRRPTSPPCARESMLV AVKGMGSQLVLRPRELNAGNVVNLSPNVDVGVS LKKGMEVTEGHKKTSVGVDVVDAGKPQCGGMM PVVQSGKEVTSELIKEATITVDKGSFSVGTTGPK PRKWNKLKRGDTSTCSVDGVKEDMSGSPDYQIT SSEIQENDAEIKIPVEIVSDEEMAFIEAAFSLASRR SSLSVLCSTSFQSISIASTRSFSSSIVCNDVEDLG KIQSPQRKKNKVSDSLLSKFRKNRGLSVTDLTST EWCEKQMEFVLTIGRPEKTNAMKAGITRHAKLEE EVSKKVEVRVQTLEDIWAVKFINFIICANQLLFDG LTREIPIVGFVQGVWMVGVIDELRMPVTGNGRKL TLVDTKTRVQPRLPTEPQRRNGRLQLMFYKRLL DNMISNNFPSEQFYVHFSMDPHNLSDEVKMNAA DSGFPAETLGEVVRYVSNVCSVLPPVDDQLLLRY ELQADNSLIGEDEFNYDEDWLKAQMHSCLEFWQ GEREAKYVPAAEKWKCQYCKFATVCPQTDTETE SQSSVQGS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
602	sp_029560_saft	#N/A	0.873385	#N/A	BEST Arabidopsis thaliana protein match is: nucleoporin-related (TAIR:AT5G20200.1); Has 61042 Blast hits to 31782 proteins in 2093 species: Archae - 202; Bacteria - 16480; Metazoa - 16017; Fungi - 12552; Plants - 1653; Viruses - 629; Other Eukaryotes - 13509 (source: NCBI BLink).	MATEASGGAGGKFRKKPVHKSGATPYDRPFTAL RPTQLVGGGGEREKNSWLSKLVDPASRLITAGA HKIFSSVFRKRLLPPPPPPQHSPGERTEIRDVQP EITTKVGSSGTKEPLVYKGNDSGGVSESSGVAD LEYILKQKTFTRSEIDHLSALLLSRTVDNLAGDEK GRSELDTSDKKVAHSPREILTNLPAIENGGERVL PAVISTPVVRSKVVEDVVASPAELAKAYMGSRPT KASSSVLGPKSDMAREDLSAPYNPFFKQKSPAV SVIQRPTSEIGASANGYMTPRYRGRSAIYSMGRT PYSRHHPTDIFKSAEMAFDTHGASSSSSSHNTTE TQQLSGSKKQVLKRRSSALESDIGFVGPIRRIRQ KPNLLHHKILSLPASGVPHIKHGYGALTEATEVPS SIDWSSNHRSSSKALAENGVHSTSNGYVPSQSI ETAQKIFQHLDKFSSKGKSPEKKCEIMGGSSAAK MTTDMREGLAHKSMSTESLKVLQTIQDSRKENG LMDSSNAHDHSLKTSPRVEGGTTQRIVIPSVEDG AEAGSSKKALEYMNHRSSFDQQQKKRSFQMSA HEDYLELDDDVHTNGVASAPMIKGVECNKLAESS AAFSDNPQSAAVSAVKPIAPEIKPSLNSLGSNKEE SLPDGSAAQAFSGFTIPTGLLTSSTTSTSKQAAV SSQFTAEINKFSSNSAEVVPISFYPSSSVVESTAA KDIKPELSISSTEAANKADELDKTAKGNTLHFGDV SKKMEGSVSSSGKSLFSFETQPNGATLTNVPST VSVSRVTSSPVPISSSSNDQRLAPVSVSSSPSLSI SNGNAVSTLSDTSAGTVINTATTGGMPIFGSNTS QFSTSIAPAVPFGSNLQFGSPSVPFTPAAATSAS ESVVNSGAKLSDSAFGNATSGQFSAKPTETASV GNNIFGFSASTLKSTSSETPPVLQAAIFSSQASSS AVASTTSIAGLSSQSAPSASSHVFGFSAPSSAYL STGANPFTTGSGSASSVFGLSSVTAPVSSSSNTT SNIFGLQSSKPAVTGSTFGASGFQFGLKASEASA TSNSTPIVFGATSSSSAAKNSTPLVFGAASSSSA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
603	sp_029850_esmd	-0.9910432	-0.9367716	-0.9801183	maternal effect embryo arrest 12 (MEE12); CONTAINS InterPro DOMAIN/s: Transcription initiation factor Rrn7 (InterPro:IPR021752); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G01335.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MAETDPSKIMCSNCDYLGLYVDDDGRYYCTNCN SQAVDIIDQANADEDFLDKGNGASIYNAANKRSA PLYSADPDRDYLGPTEPADFGKTPKGVYYKDED YYNTIRMRHILGVQLMIQFQCEALVEKCGVTPLIC GIAASIWLRFVATTRVFDDDWADKTVLDSETVSP SVFNFRFMIFQGEKRYQVGKHCINEPRNMHNER MVMIWSRSLKKTIPLSCSLAVSYLACHIAKEPVLP TDIIKWSIEGKLPYFGAFPKIAESIIEIEKYDHHVSVI CPLKTSTMFRPIQLPLHKLESMAASIAQSVGLNLP PVNFYRIAHRYLKQLSLPVEKILPYACRIQEWLMP PDLWLSASEEKLPSRVCVMSILVVAIRILYNLNGF GKWESTLSAPTVFSSRDDQTASSGKSNANKGAE KASKNLGGESGANLSSVHKFEFDSTELLRNLERR YDDLSDKFEYSKDLPTYLQYGKDVVFAGVEPLYK DPTFIEELWDFYEKNREDIKTSEAFETESNQKRS RDNDKIFSTPSKDNKKHRDEGITSSSSADSGFQC IYNSPKWDQGADNTQQSSQADQAPATVEKGEA SLEDIALKYLKIDMEEHRFCYIPPRVEIKRRDHLHY SRKKADGSLAFVAHADYYILLRACAKVAQVDIRIM HVSVLSLERRLSWIEKWIDNSFLEKCSTGSVESE DEDGLEDEYVNLSNLNI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MTTDLTRNGPVERDIEQATTALCKKGATCLKTGRR
						GKPKFCPFRLSNDESVLIWFSGKEEKHLKLSHVS
						RIISGQRTPIFQRYPRPEKEYQSFSLIYNDRSLDLI
						CKDKDEAEVWFSGLKALISRSHQRKWRTESRSD
						SIPSEANSPRTYTRRSSPLHSPFASGDGLQKDSS
						DHLRLQSPYGSPPKNALDTAFSDVMFTAPPKGF
						FPSDSATMSVHSISSGSDSVYNHMKGMGPDTFR
						VSLSSAVSSSSQGSGLDEGDALGDVFIWGEGTG
					Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain; FUNCTIONS IN: chromatin binding, zinc ion binding, Ran GTPase binding; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Regulator of chromosome condensation, RCC1 (InterPro:IPR000408), Disease resistance/zinc finger/chromosome condensation-like region (InterPro:IPR013591), Pleckstrin homology (InterPro:IPR001849), Zinc finger, FYVE-type (InterPro:IPR000306), Regulator of chromosome condensation/beta-lactamase-inhibitor protein II (InterPro:IPR009091), Zinc finger, FYVE-related (InterPro:IPR017455), Pleckstrin homology-type (InterPro:IPR011993), Zinc finger, FYVE/PHD-type (InterPro:IPR011011); BEST Arabidopsis thaliana protein match is: Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain (TAIR:AT5G12350.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	VQNIACGGRHAALVTKQGEIFSWGEETGGRLGH
						GVDADVSHPKLIDALNTTNIELVACGEHHTCAVTL
						SGELYTWGDGAFNFGLLGHGNDVSHWVPKRVN
						GPLEGIHVSYISCGPWHTAVVTSAGQLFTFGDGT
						FGVLGHGDRKSVSMPREVDSLKGLRTVRAACGV
						WHTAAVVEVMVGSSGSSNCSSGKLFTWGDGDK
						GRLGHGDKEAKLVPTCVAALVEPTFCQVACGHSI
						TVALTTSGHVYTMGSPVYGQLGNPHSDGKLPTR
						VEGKILKCFVEEIACGAYHVAALTSRTEVYTWGK
						GANGRLGHGDTEDRNGPTLVGALKDKQVRSIAC
						GTNFTAAICLHKWVSGVDQSMCSGCRLPFNFKR
						KRHNCYNCGLLFCHSCSSKKSFKASMAPNPNKP
						YRVCDNCLAKLRKAVETGASTRSSASRRGIVNQ
						GSESNEREDKLDSGSRGQLARYSSMDSSEGRS
						KKNKKYEYDSSRVSPIPNGGSQWGANISRSFNP
						MFGSSKKFFSASVPGSRIVSRATSPISRRPSPPR
						STTPTPTLSGLMSPKVTVDDAKRANDGLSQEVIR
						LRAQVETLTRKAQVQEVELERTNKQLKEAIAAG
						EETAKGKAAKEVIKSLTAQLKEMAERLPVGATRN
						NKSPSFPSLSSSTPTFADISNSSIEQFNDQIIPSLD
						SFGFNGLVLSNGSATPSSRSSVQSRIVNSEATRN
						GNRTREGEPRPENEWVEQDEPGVYITFVSLPGG
604	sp_031040_frad	-0.8581166	-0.8581138	#N/A		IKDLKRVRFSRKRFSEKQAEQWWAENRARVYEH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
605	sp_031110_aqmg	-0.8011336	#N/A	#N/A	Heat shock protein DnaJ with tetratricopeptide repeat.; FUNCTIONS IN: heat shock protein binding, binding; INVOLVED IN: protein folding; LOCATED IN: chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide TPR-1 (InterPro:IPR001440), Tetratricopeptide-like helical (InterPro:IPR011990), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734), Heat shock protein DnaJ, conserved site (InterPro:IPR018253); BEST Arabidopsis thaliana protein match is: Heat shock protein DnaJ with tetratricopeptide repeat (TAIR:AT2G41520.1); Has 27844 Blast hits to 24585 proteins in 3132 species: Archae - 222; Bacteria - 9738; Metazoa - 5941; Fungi - 3019; Plants - 3250; Viruses - 11; Other Eukaryotes - 5663 (source: NCBI BLink).	MNSYTRQSGFVNSSEMNAKTEFSSPSTSNLGSEI PTFSSISGLSKPRLMKSRKSQNAKSMGSSETRRI DPGFNPFRPVLDNLAGVSSRLESEDDNGAFMNS SNSSKHLNLKVTSKVEKVGDELIDEIRKLRIENQNI ARQSKGVSNDLPSELPDKIKNLNLGGSGRENSR VSGVVVDNVSGSFGTSMESELPIDLERLNLKDTG EMGNPNVVFGASGSNSFVFGSTRTMKDSTDPSI ASTLPNRIKNLNIADPDKNSGTKKEMNINQNNRT ANRTESGEVDTMILDEMAKLEILGDSQESSDHPR TEQVGSMSKEDSRIPCQPKSRNQINEHDEHFIGP QVAETQSDKGFGFYDASTSSFSSAGVHFQPVRH EFQAPCTDGAEKKEGFTFSSKQDGLRIPHVELTP DSKGNLFTGSKVQTGFSTKKGLVKDAKSKGRKA KVKEPKSVLLSFANNFTSKENNMQPEPSESYSP MDVSPYRETEYDEQLSREHSLTSENQTSLGSQT SMVSNDAIDEDLIRATESMDINESNAVKHDEADD QVQVDCVSAVESESFMSASENLDENGENEVESN SEECGYFSAASFASHGESSFTFAASSSNQDQAS AAMRFNKKKIRMKGGSDIYIPTRASVLSDAQFSPI ARVSVLPSQVVQSDVLPTAPPKEGKKITSPTTSM LLQSTDVCVERKILEASDGLQKAQKASEFMDRS SGLLLQRTLVDAESALGVIAEALIISPYSEKLLESK AEALLMLRRYEEVIQLCGQTLETAEKNSPLLSAD GQSLYLEASEQLLKSYSFRIWRCCDMFKSNFYL GKLEEAVEFLGKQECWMSMVAKFGNETLESFIP QLSTARELINHKSAGNGAFQAGKHKEAVEHYTA ALSFNVESRPFSAVCFANRAAAYCALGQITDAIA DCSLAIALDENYLKAISRRATLFEMIRDYEQAAKD MLKICDEAAALREMHRLKSRKRWTTMFKGSDLLM CGI
606	sp_031200_qoqf	-0.662321	#N/A	#N/A	uncharacterized protein	

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
607	sp_031440_awod	-0.906094	-0.850777	-0.89138	HIS triad family protein 3 (HIT3); FUNCTIONS IN: protein kinase C binding, adenylylsulfatase activity, zinc ion binding; INVOLVED IN: sulfur metabolic process, purine ribonucleotide metabolic process; LOCATED IN: peroxisome, plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histidine triad-like motif (InterPro:IPR011146), Histidine triad, conserved site (InterPro:IPR019808), Histidine triad (HIT) protein (InterPro:IPR001310), Histidine triad motif (InterPro:IPR011151); BEST Arabidopsis thaliana protein match is: HISTIDINE TRIAD NUCLEOTIDE-BINDING 2 (TAIR:AT1G31160.1); Has 7957 Blast hits to 7956 proteins in 2592 species: Archae - 148; Bacteria - 4849; Metazoa - 343; Fungi - 139; Plants - 127; Viruses - 0; Other Eukaryotes - 2351 (source: NCBI BLink).	MASEKEAALAAASAPADSPTIFDKIINKEIPANIVF EDEQVLAFRDINPQAPTHIIIIPKVKDGLTGLSQAE ERHYEILGRLLHTAKLVAKQEGLDDGFRIVINDGP KGCQSVYHIHVHLLGGRQMNWPAG
608	sp_031490_gamj	-0.848552	#N/A	#N/A	uncharacterized protein	MFYSVPIYDMSLVNGQRDCLCSRPASYGDDQAA QHHELALPTRD
609	sp_032880_gxac	-0.975239	-0.990271	-0.981105	uncharacterized protein	MTKQTAVGRVVAAGVAQCKTGWERETWLLTKK CRGCGKSPR
610	sp_033170_gmxs	-0.90928	-0.962338	-0.892787	ATUBC2-1; CONTAINS InterPro DOMAIN/s: Ubiquitin-conjugating enzyme/RWD-like (InterPro:IPR016135), Ubiquitin-conjugating enzyme, E2 (InterPro:IPR000608); BEST Arabidopsis thaliana protein match is: ubiquitin-conjugating enzyme 18 (TAIR:AT5G42990.1); Has 9046 Blast hits to 9044 proteins in 385 species: Archae - 0; Bacteria - 0; Metazoa - 3956; Fungi - 2002; Plants - 1701; Viruses - 20; Other Eukaryotes - 1367 (source: NCBI BLink).	MTSSSTQSRKVLSKIAQNRLQKELAEWQLNPPP GLKHKITDNLQRWVIDVSGAPGTLYANETYQLQV DFPEHYPMEAPQVIFVPPAPLHPHIYSNGHICLDI LYDSWSPAMTVNSICISILSMLSSSTSKQRPPDN DRYVKNCRNGRSPKETRWWFHDDKA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
611	sp_033250_jpic	-0.992948	-0.977115	-0.966385	phytochrome-associated protein 2 (PAP2); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: response to auxin stimulus, regulation of translation; LOCATED IN: intracellular, nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aux/IAA-ARF-dimerisation (InterPro:IPR011525), AUX/IAA protein (InterPro:IPR003311); BEST Arabidopsis thaliana protein match is: indoleacetic acid-induced protein 8 (TAIR:AT2G22670.4); Has 2237 Blast hits to 2234 proteins in 81 species: Archae - 0; Bacteria - 0; Metazoa - 1; Fungi - 0; Plants - 2235; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	MSRPFQHDYIGFSEVSSMENSVKSPNFSLKTELS LGLPGSESPERKPGSYGVTIFGKDFEGKSQNGY SICSPKSSASGAKRGFSDAIDGCEKWVFSINGNS DPNLSKDGAKNGKNFTGVENNNVPSSPKPVQDK QPQLSAANELASAPAAKTQVVGWPPIRSFRKNT LASSSPKNNGHTEEKSVCLYWKVSMDGAPYLRK VDLQTYRNYMELSSALEKMFSCFITGQYCSNGH GKQDADLQNAEYVLTYEDKDGDWMLVGDVPWR MFTENCRKLRIMKGTEAIGLAPNAAEKSNTEA
612	sp_033500_ihfk	-0.870366	#N/A	#N/A	ACCUMULATION OF PHOTOSYSTEM ONE 4 (APO4); INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF794, plant (InterPro:IPR008512); BEST Arabidopsis thaliana protein match is: Arabidopsis thaliana protein of unknown function (DUF794) (TAIR:AT5G57930.2); Has 163 Blast hits to 121 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 163; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MALFRHRNWCKILVDNIEINEIGRRWYATNTKVD WDKLRPMIKKRIENRARGYPIKSMIPVAEQVLQA RSDLIRGVSALLKVLPVVSCKYCPEVHVGESGHLI RTCCGHRRLNKNQDHQWIRASLNDILVPVETFHL DNMVQDVIRHEQRFDFERVPAIVELCYQAGVIPH DLGLQSLDYLSNSTVDFALTGPDSLSASDLRTIG NRTLRAWETLRHGVQKLLLVYSAKVCKYCSEVH VGPSGHKARLCGVFKYQNFRGNHFWEKAGVDD LVPPKIVWCRRPQDPPVLLHEDRNYYGHAPAVV DLCMKAGTVPPVKYLCMMKAQGLRGPSPDKVL G

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
613	sp_033650_qdon	-0.847611	#N/A	-0.87737	BRI1 suppressor 1 (BSU1)-like 1 (BSL1); FUNCTIONS IN: protein binding, protein serine/threonine phosphatase activity; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Serine/threonine protein phosphatase, BSU1 (InterPro:IPR012391), Metallophosphoesterase (InterPro:IPR004843), Galactose oxidase/kelch, beta-propeller (InterPro:IPR011043), Kelch-type beta propeller (InterPro:IPR015915), Serine/threonine-specific protein phosphatase/bis(5-nucleosyl)-tetraphosphatase (InterPro:IPR006186); BEST Arabidopsis thaliana protein match is: Serine/threonine protein phosphatase family protein (TAIR:AT1G03445.1); Has 11450 Blast hits to 9180 proteins in 548 species: Archae - 80; Bacteria - 322; Metazoa - 3883; Fungi - 1977; Plants - 2720; Viruses - 6; Other Eukaryotes - 2462 (source: NCBI BLink).	MASKPWLQPAPSYTQLESFWDSDEDAPGPRCG HTLTAVAANKSQGPRLILFGGATAIEGGASSSAP GIRLAGVTNSVHSYDVLTKKWTRIRPAGDPPSPR AAHAAAAVGTMVVFQGGIGPAGHSTDDLYVLDL TNDKYKWHRVVVQGQGPGPRYGHVMDLVAQR YLVTVSGNDGKRVLSDSWVLDTAQKPYVWQRL NPEGDRPSARMYATASARSDGMFLLCGGRDST GAPLADAYGLLMHRNGQWEWTLAPGVSPSPRY QHAAVFVGARLHVTGGALRGGRAIEGEAAVAVL DTAAGVWLDRSGLVTTSKANKGQTENDPSLELM RRCRHASASVGVRIYMYGGLRGDVLLDDFLVAE NTVSQSDDNSLPLASERASSLTDPRINLSLTMPS ATPVTDGRHEKTLPGGVSMEKKSVDKLVEASAA EVEAVRTVWQAAQAASSPPDESPLSQGSSPIT DTPSDGSDTDGDVRLHPRAVVVAKETVGNLGG MVRQLSLDQFENESRRMIPVADQSTLGKKFTRS KSPQGLHKKVISNLLRPRNWKPPVNRKFFLDSYE VGEIC-HAAEQIFMHEPSVLQLKAPIKVFGDLHGQ FGDLMRLFDEYGFPSTAGDITYIDYLFLGDYVDR GQHSLETITLLLALKIEYPESVHLIRGNHEAADINA LFGFRLECIERMGENDGIWAWTRFNQLFNYLPLA ALIEKKIICMHGGIGRSIQAVEQIEKLERPITMDAG SIVLMDLLWSDPTENDSIEGLRPNARGPGLVTFG PDRVTEFCKKNKLQLIIRAHECVMDGFEERFAQGQ LITLFSATNYCGTANNAGAILVIGRGLVVVPKLIHP LPPPMQSPETSPERAMDDTWMQELNIQRPPTPT RGRPQPDHDRNSLAYI
614	sp_034220_totz	#N/A	-0.916571	-0.938888	RING/U-box superfamily protein; FUNCTIONS IN: zinc ion binding; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein (TAIR:AT1G17145.1).	MGNKLGKRRQVVDEKYTRPQGLYQIKDVDYKKL RKLILESKLAPCYPGGDECDCGLLEECPICFLYYP TLNRSRCCMKSICTECFLQMKNPNSTRPTQCPF CKTSNYAVEYRGGKSKEERGLLEQIEEQKVIEAKI RMRQQEMEEEEERQHKKEETSSSERDPRTENID CSTSGGVECEETAPSQDSNGTQMLRQPHTRHG RDDEFDLDLEDIMVMEAIWLSIQENGRAEYSLSN DEVPEQYTAGDHYMSPVTAPAVGSSSSPSGGLA CAIAALAERQQSTGESSTHYSEPVSRLDMLPSSS RSSKRMKKTQRKHAECSSKQICSSSMISEGCEV SPALEVAEVGTSYINSDDSEGEGVILPPPPPPPM YRRWNSKRGKNIQRKYSFERASFKNLHDRRVVA EDWEFDHGSEVAEAGTTYASSDESNEDIQAAVL PPPPPPPPLLPPYVVGSGSGCSDHGEYLLPPP YVTDSGFQSPIPGPIVPENFEEQMMLAMAVSLAE ARARTTAQGVAWH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
615	sp_034910_oips	-0.847785	-0.883142	-0.9461185	DC1 domain-containing protein; FUNCTIONS IN: oxidoreductase activity, antioxidant activity; INVOLVED IN: response to cadmium ion, pollen tube growth, pollen tube guidance; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen (InterPro:IPR000866), Thioredoxin-like (InterPro:IPR017936), Thioredoxin-like fold (InterPro:IPR012336), C1-like (InterPro:IPR011424), Thioredoxin, conserved site (InterPro:IPR017937); BEST Arabidopsis thaliana protein match is: protein kinase C-like zinc finger protein (TAIR:AT4G31240.2); Has 6688 Blast hits to 3903 proteins in 794 species: Archae - 4; Bacteria - 4185; Metazoa - 634; Fungi - 4; Plants - 553; Viruses - 0; Other Eukaryotes - 1308 (source: NCBI BLink).	MDSARKRKSADSATVLKKTGDIVDDFWPLKLKD PFENLSRFQENGKCYNRFFYSFLRVSKDGEYGD DIKKMVSELPDMEIGDPFDVLSLLSFKGQSNYLV RNNNDNDQVELNKDTFHGKYYMVCWFHVPIHCL DGDALFIQSAANICYEMFDLFDDFEIVLGLKMNPK YNNQRVFSHFLSAFPPSCLVVPFDDPKRSDYICC YLEVRIRCECFLVDGNSIALFSGEPDFEMGERVS ASRLCSNYSAKDLEIISLPKITRLRSIINNTVVRRIN EEGVQEEMSFLRLNEKKVVGLYLWTEGIRCITDE SFIDTLDMVNKQCRDILVADALQHHMQQTWEHC GCIPGQLVIVGPNTEYVDLYGRDILGKYGSVAYPF TREGLISNGVKLLRELTLESLLTSESCNYVLKDGF TQINVASLQPRDILLYFGSEDGSLIKSLQNLYSEV KLENFNFEVIYVRRNNTSNYCNNMDIGLPCLVRP FEADHSSSLGKKLFPHVVPVDTFVRFGNDGHVC SVGGNNLREDVCSVGGNSLREDVIFMLERLGIME DYIRYCC
616	sp_034990_pexn	-0.945451	-0.9629328	-0.9875775	unknown protein; Has 122 Blast hits to 122 proteins in 57 species: Archae - 0; Bacteria - 0; Metazoa - 46; Fungi - 30; Plants - 41; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MEAQTLSSSWQFSCDLEVNFESDEKAYIVYAALN VDKELQPDKVKRVMSVSDGKLAVHFQAVEARFL RASFSAFVDVLRLATRTIEEFSPGMEI
617	sp_035630_neiu	-0.8648635	#N/A	#N/A	uncharacterized protein	MENKINNDDREGGVNSLSEPNMSEALCICCAL WMEKIGSVMEGMKVRMKW
618	sp_035650_rkws	#N/A	-0.877447	-0.8666347	phloem protein 2-A15 (PP2-A15); FUNCTIONS IN: carbohydrate binding; INVOLVED IN: N-terminal protein myristoylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364); BEST Arabidopsis thaliana protein match is: phloem protein 2-A12 (TAIR:AT1G12710.1); Has 510 Blast hits to 504 proteins in 28 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 510; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGASLSNSAEGGANGGENGPSLGDIPEAGIACV FMYLTPPEICNLARLNRAFRGAASSDNVWESKLP PNYQDLLDLLSPERYQGFSKKDIFSLLARSISFED DTKQVCVDKVTGRVCLSVSARAMSITGIEDRRY WSWLPTEESRFQVIAYLQQVWWFEIDGVVKFPL PPDIYTLTFRLHLGKFSKRLGRRVSNFEHTHGWD IKPVRFELTTWDNQHATSECYLDDTEPDEMKEN LKRGCWIEYKVGEFIVRDSDSVPEVRFSMKQIDC THSKGGLCVDSISIVPSDLREERRRRLLSN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
619	sp_035830_cyaw	#N/A	#N/A	-0.866331	XIK; FUNCTIONS IN: motor activity; INVOLVED IN: in 11 processes; LOCATED IN: endomembrane system, myosin complex; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: seedling growth; CONTAINS InterPro DOMAIN/s: Dil domain (InterPro:IPR018444), Dilute (InterPro:IPR002710), Myosin, N-terminal, SH3-like (InterPro:IPR004009), Myosin head, motor domain (InterPro:IPR001609), IQ calmodulin-binding region (InterPro:IPR000048); BEST Arabidopsis thaliana protein match is: Myosin family protein with Dil domain (TAIR:AT1G54560.1); Has 40570 Blast hits to 26875 proteins in 2480 species: Archae - 466; Bacteria - 5527; Metazoa - 21427; Fungi - 3168; Plants - 2202; Viruses - 135; Other Eukaryotes - 7645 (source: NCBI BLink).	MATETVVGSKVWVEDPELAWIDGEVTKTSGQE LHVKTTNGKSIVTEVSKVYPKDEEAPAGGVDDMT KLAYLHEPGVLSNLAMRYSINEIYTYTGNILIAVNP FQRLPHLYDTHMMGQYKGAGFGELSPHVFAIAD ASYRAMINEGKSNSILVSGESGAGKTETTKMLMR YLAYLGGRSGVEGRTVEQQVLESNPVLEAFGNA KTLRNNNSSRFGKFVEIQFDKHGKISGAAIRTYLL ERSRVCQVSDPERNYHCFYLLCAAPQEVRDKYK LDSPDKFHYLNQTNCYKLDGVDDSEEYLATRRA MDIVGISAEDQEAIFRVVAAILHLGNVNFAKGTEV DSSVLKDEKSRFHLAMTAELLRCDIKSLEDALIKR VMVTPEEVITRTLDAHSAQISRDAFAKTIYSRLFD WLVDKINNSIGQDPNSKSIIGVLDIYGFESFKLNSF EQFCINFTNEKLQQHFNQHVFKMEQEEYTKEEIN WSYIEFVDNQDVLDMIEKKPGGVIALLDEACMFP KSTHETFAQKLYQTYVKHKRFQKPKLSRTSFTIC HYAGDVVYQADQFLDKNKDYVIAEHQDLLQASK CSFVASLFPPLPEETSKSSKFSSIGSRFKLQLQQL METLNSTEPHYIRCVKPNNLLKPAIFENVNILNQL RCGGVLEAIRISCAGYPTRRTFEDFLLRFGVLAPE VLDGKCDDKIVCQKILDKMGLKGYQIGKTKVFLR AGQMAELDTRRAEVLGNAARKIQRHGRTYIARKE FISYRNAAIQIQSWCRGVLACKVHEELCRQASAIR IQKNFRCYKARQSYLRVRLSAVTVQTGLRAMHA RNMFRYKKETKAAIIIQTRQRCHMAYSYYQNLQK AALVTQCGWRRRVARKELRQLKMAARETGHLK ELKDKLEKRVEDLTWRLQFEGQLRKNLEEEKGQ EIAKLQEALQAMKMQVEEANARITKEREAAKKAV EDAPPVIKEVPVMVQDTEKVESLTIEIENLKALLE KEKEAAEVTRKSVSDTEERNAELVKKLEGSDDTI NQLKHFVQRMEEKLSNAESESQVLRQQGLEMT QKLSSAESEIQDMAEKLKKAESELQALRQQALTM	
620	sp_037040_wdpn	#N/A	-0.865378	#N/A	Regulator of chromosome condensation (RCC1) family protein; CONTAINS InterPro DOMAIN/s: Regulator of chromosome condensation/beta-lactamase-inhibitor protein II (InterPro:IPR009091), Regulator of chromosome condensation, RCC1 (InterPro:IPR000408); BEST Arabidopsis thaliana protein match is: Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain (TAIR:AT5G19420.1); Has 19815 Blast hits to 6041 proteins in 472 species: Archae - 85; Bacteria - 2553; Metazoa - 7293; Fungi - 979; Plants - 2476; Viruses - 2; Other Eukaryotes - 6427 (source: NCBI BLink).	MEERERVLKNQKKTEQEEHEQEKDIEPKQQELW SWGAGTEGQLGTGNLKDELTPQLLINLSNSLSSF APISILSCGGAHVIALTSGGKVLSWGRGSSGQLG HDQIMPNCSFPKLVASLSSFNISHVSAGWNHSGF VSDSGCLFTCGDGSFGQLGHGDFKSYSSPIEVS FFNSKHVEQVACGMRHSLVLSKGNSGDKVMYG FGSGKRGQLGISEDKIKLVSVPVVCSGFENAEVV DILANGDQSAALSADGRLYTWGRGFSSSSSNVCI PQSIPSAFSFEQAALGWNHALVLTSEGEVFMLG GSHHGVLSEPQKMTRHNNSDEPSLSKVCGLV GKKVVHIAAGAEHSAIVTENGQVYTWGWGEHG QLGLDSTSDHTAPQLVLPEDRSLISRVYCGSGFT FALKTHT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
621	sp_037140_nteo	-0.863193	#N/A	-0.880048	uncharacterized protein	MSQLTNQEKGDVDADTWLNELYQQRVKLHQIQD ETKASSKAVKPSFVSRSALASKRKKEDEAKAPSK KPKLKITKKVKKEKETVEKFHDKFHDGPAKELTLA VAGSKGAINVLSAFDEDKVTPAGKEREIEVSTNG GGTSETPAKQLVKAKPLGDLPEVVGDKVPSFVF DRVRQTGEASLEQKWDDEEVGEFPSTKRFDNK SQAKIVEGIMNVVPKEYREFVQGPDDAQLGAIQA TLVDLLIKMDGAKKWHREVQRDLKKQVTKAAFV QIEEKINEQILAIDTITGELDKALAAKAELKAKLEAQ SSMEARVVELEKDLAAARADVAFLSADAKSAYIK GEEDTKKATHMAWDFVT
622	sp_037170_ikrj	-0.88696	-0.847604	#N/A	C2H2 and C2HC zinc fingers superfamily protein; FUNCTIONS IN: zinc ion binding, nucleic acid binding; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, U1-C type (InterPro:IPR013085), Zinc finger, U1-type (InterPro:IPR003604), Zinc finger, C2H2-type matrin (InterPro:IPR000690); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MPRYYCDYCDTYLTHDSPSVRKQHNEGYKHKAN VRIYYQQFEEQQTQSLIDQRIKEHLSGQGAFQIG AAYAAWRPGFRPPLLPAPLPIPGSLPPPNAQMA PGIRLPVFPRPPGPPGFPVYGAPPGQIAGQPGA PGPPGPPGAPVQANG
623	sp_037510_fxuf	-0.89532	-0.951936	-0.893447	Nucleotide-sugar transporter family protein; FUNCTIONS IN: nucleotide-sugar transmembrane transporter activity, sugar:hydrogen symporter activity; INVOLVED IN: carbohydrate transport, nucleotide-sugar transport; LOCATED IN: integral to membrane, Golgi membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Nucleotide-sugar transporter (InterPro:IPR007271), UDP/CMP-sugar transporter (InterPro:IPR021189); BEST Arabidopsis thaliana protein match is: Nucleotide-sugar transporter family protein (TAIR:AT4G35335.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	REVIVTSSVLACEVAKVICAFTVLIFNGSLPKLCKD WSLLRALTFSGLPAAIYALQNSLLQISYKNLDSLT FSMLNQTKLFFTAFFGYMILRQKQSIQQIGALFLLI LAAVLLSIGEGTSKASKSRNPEEVLFYGIVPVLIAS VLSGLASSLCQWASQVKKHSSYLMTIEMSIFGSL CLLTSLSQSPDGTAMRKHGFFQGWTPLTLVPVV ANALGGILVGLVTTYAGGVRKGFVIVSALLVTALL QFICDGIPPSPYCLVALPLVITSVSIYQKYPYRVKK KEA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
624	sp_037550_dcrd	-0.8807755	-0.388326	#N/A	Hypoxia-responsive family protein; CONTAINS InterPro DOMAIN/s: Hypoxia induced protein, conserved region (InterPro:IPR007667); BEST Arabidopsis thaliana protein match is: Hypoxia-responsive family protein (TAIR:AT5G27760.1); Has 144 Blast hits to 144 proteins in 38 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 45; Plants - 99; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAKESKTTMESVRSWVFQHKLRTVGCLWLSGIN GSIAYNWSRPQMKTSVRLIHARKEELLNYFCRF
625	sp_037790_xmku	#N/A	-0.868365	#N/A	uncharacterized protein	MEAVVSELMASPLPGSLDRERGKDIVCGLSVVS PGDDIECDNKMDGLLASSALLFQRSRPRRGRKK RVALADRGRGWFTSEIKKFSYNLVEFVNHNGGL LSGRGNLKGLSLNVMIKCGDGSNCNMHQCKAVA QLNGCVHKQEKVMKFAADCRDGVHNSAVDRSR LANWMSKEREKMKVASHLGEEDQHTETGTVFS RISELIGVDKLLFGGEDGDGWGELLSWERRRMD IDIQKDTHSCGVPMLLAIKDFAKGYESLRITDLEE
626	sp_038000_ftpx	-0.9837663	-0.954471	-0.983318	Mitochondrial substrate carrier family protein; FUNCTIONS IN: binding; INVOLVED IN: transport, mitochondrial transport, transmembrane transport; LOCATED IN: mitochondrial inner membrane, membrane; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Mitochondrial carrier protein (InterPro:IPR002067), Mitochondrial substrate carrier (InterPro:IPR001993), Mitochondrial substrate/solute carrier (InterPro:IPR018108); BEST Arabidopsis thaliana protein match is: Mitochondrial substrate carrier family protein (TAIR:AT2G30160.1); Has 27351 Blast hits to 14015 proteins in 457 species: Archae - 0; Bacteria - 6; Metazoa - 11451; Fungi - 7735; Plants - 5292; Viruses - 0; Other Eukaryotes - 2867 (source: NCBI BLink).	ARRLLLTQDILSPFNLERDRIKELVNGANRKKL MATTDATPKFHNHELRPLSSLEEQQQQNLTMDP SPSHTHHDGLLFWQFMVAGSIAGCAEHMAMFPV DTIKTRMQALGSCPIKTAGVRTALRSILQSDGPA GLYRGIAAMGLGAGPAHAVYFSMYEFFKKSFSG GNPNNHYAHAMSGVCATVMSDAVLTPMDMVKQ RLQLSGGNGGAYRGVWDCVKLVLKNEGFGAFY ASYRTTVIMNAPFTAVHFATYEAAKKGLTEISPDT ADDERLVVHATAGAAAGALAAVITTPLDVVKTQL QCQGVCGCDRFSGGSIRDVLKTIVKKDGYRGLM RGWAPRMMFHAPAAAICWSTYEAAKTFFQELND QKNSSSVT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
627	sp_038170_jmqo	-0.8870245	#N/A	-0.9884585	Prefoldin chaperone subunit family protein; FUNCTIONS IN: unfolded protein binding; INVOLVED IN: protein folding; LOCATED IN: prefoldin complex; EXPRESSED IN: pollen tube; CONTAINS InterPro DOMAIN/s: Prefoldin alpha-like (InterPro:IPR004127), Prefoldin (InterPro:IPR009053); Has 429 Blast hits to 422 proteins in 142 species: Archae - 27; Bacteria - 8; Metazoa - 215; Fungi - 17; Plants - 37; Viruses - 4; Other Eukaryotes - 121 (source: NCBI BLink).	MATARGKGTVTSLSSTFPAEEARKAADLVNHTIN EKRSELDNLRSFVADNTNLINLVQRLPDELHHHIQ VPFGKAAFFPGRLVHTNEFLVLLGEGYFADRTAK QTVDILKRRGQTLESHVDSVKAVIKDLKAEASFFY RTAEEAAEGLVEIREELSDDEDSEGSQEKEDAAS EPEVGIGSSAAVLDDEEYARIMSRLDELEKEELE EEEEDEDTEDHRDVTNGEDFNNSEQSEVQHHTK VLSDSISLPNDADKSIHTYLNLQPNHKDDCSRGK DILQKFVERPMEEDNVPSVVKTSFPSESRSKSEV YVPAPKPEVGNQLSRQVKTDSLKAFTGSIIERTD NLQISPDREPINQPSNSQPSKPVSRFKMQRRGV
628	sp_039020_guqe	#N/A	-0.8953112	#N/A	alfin-like 6 (AL6); CONTAINS InterPro DOMAIN/s: Zinc finger, PHD-type, conserved site (InterPro:IPR019786), Zinc finger, PHD-type (InterPro:IPR001965), Protein of unknown function DUF3594 (InterPro:IPR021998), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: alfin-like 7 (TAIR:AT1G14510.1); Has 1820 Blast hits to 1777 proteins in 193 species: Archae - 0; Bacteria - 0; Metazoa - 798; Fungi - 350; Plants - 563; Viruses - 0; Other Eukaryotes - 109 (source: NCBI BLink).	MEGMPQNTPRTVDEVFNDFKGRFVGLIKALTTD VDKFYQQCDPEKENLCLYGILFNETWEVNLPVEE VPPELPEPALGINFARDGMQEKDWLSLVAVHSD SWLLAVAFYFGARFGFTKNERRKLFQMINDMPT VFEILNGNAKQPKENPGPHNSSSKNKSGVKPSR QPENHSKPMKMPTPPKEAESGGEEEEEDDEQG ATLCGACGDNYASDEFWICCDVCEKWFHGKCV RITPAKAEHIKQYKCPGCSTKRARV
629	sp_039160_anky	#N/A	-0.8683196	#N/A	uncharacterized protein	MRNNRKLFSISFTTLAEIQWYTLPQPLYKQYYK KKQQKYKTTVQITTNKGVEIKLGSKLCKNMAKAT QSAAEKAVRKLQENFNFDVADINYTTCMHQRKK RRIAEIKLTREEMNFIDLKRYYFQECAENRSLKEQ LVKLLNRDNKGSG
630	sp_039300_htyn	#N/A	-0.8830195	#N/A	uncharacterized protein	MTYRKRVRATAKTDAIGSASKSLPSRTSSLSEVA SQPTKKHATQSQPSQVPKSVIFGPFSPPGTTTS MVAPPCSSTAKSAQQAIWKTPPTGSASNSLEAS TKNKIPPKITSKISIQPNIPPSKPLKYPSTSSGQIPP NMPPHVPPSSHTKTSAGDSQSQQKSVQEPKEN STKSKSLREYPDWRKNGAKM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
631	sp_039570_npik	#N/A	-0.8567776	-0.9273384	Trimeric LpxA-like enzyme; FUNCTIONS IN: binding, transferase activity, translation initiation factor activity, nucleotidyltransferase activity; INVOLVED IN: biosynthetic process, regulation of translational initiation; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Trimeric LpxA-like (InterPro:IPR011004), eIF4-gamma/eIF5/eIF2-epsilon (InterPro:IPR003307), Armadillo-type fold (InterPro:IPR016024), Nucleotidyl transferase (InterPro:IPR005835); BEST Arabidopsis thaliana protein match is: Trimeric LpxA-like enzyme (TAIR:AT4G18300.1); Has 8121 Blast hits to 7961 proteins in 1548 species: Archae - 574; Bacteria - 3735; Metazoa - 1296; Fungi - 469; Plants - 387; Viruses - 0; Other Eukaryotes - 1660 (source: NCBI BLink).	MVRRTSRAAEDADDLTHVPLQAILLADSFATKFR PITLERPKVLLPLVGIPMIDYSLEWLESAGVEEVF VFCCAHSKQVIQHLAKSQWLKNPDMAVTTIESH NSISAGDAMRVIYERNVIHGDFVLVTGDTVSNMS LAEALQEHKDRRKKDSNAVMTMVVKKSKPSPIT HQSRLGTDELFVAINPDTKQLLSYEDNANPIDKSL LSDNSSISLRNNMQDCYIDICAPEVLSLFTDNFDY QHLRRDFLKGLLVDEITDYKIYTHEIHSSYAAKVE NFRSYDTISKDIIHRWTYPLVPDVQYFGSSAVKLG RQGVYSASDIGLSSNAQIGPFTLIGKGTSVGSNS KVSNSIIGEGCKIGSNVSIEGSHIWNNVTIEDGCK LKHAILCDGVVMKARAVLEPGVILSFKVVIGKDFV VPAYSKVSLYEQPIFQDSDEELEYADSSNGMAEI SSISASTGMPDGELVSRAAEASEVEADGAGYIW SVFEGREEEWRHSVAPISEDKLAELLQAADDVLD VPHVENHQHSEELVAPSIDPENLTDDNIDEFEKE VEATFQRAVQENIKDDNVVLEVNSLRLSYNMAAA DCAGALFYSVMKLAVETPHSSISELLHNVTRLIRK WAKLLKYYLSSIDEEIEVILKFEEMCTESTKELSPL FTKVLQFLYEDVDLIQEEAILSWESEKMGADEVD RVFLKQAEPFIQWLKTAQEEDSEEEE	
632	sp_039600_nsgq	#N/A	-0.835631	#N/A	zinc finger (C2H2 type) family protein; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: intracellular; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880); BEST Arabidopsis thaliana protein match is: zinc finger (C2H2 type) family protein (TAIR:AT5G16470.1); Has 86 Blast hits to 86 proteins in 31 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 77; Viruses - 0; Other Eukaryotes - 9 (source: NCBI BLink).	MTGKAKPKKHTAKELAAKLDAATTNRGGGKAGV ADRTGAVKGGHAKMECPLCKVTAPDVKSMKIHH EAKHPKIPFDENKISNLHASTPAAAPSTSVADPNK PKPGIRGSLKK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
633	sp_039670_occm	-0.853053	-0.871838	#N/A	Peroxidase superfamily protein; FUNCTIONS IN: peroxidase activity, heme binding; INVOLVED IN: oxidation reducion, response to oxidative stress; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: Haem peroxidase (InterPro:IPR010255), Plant peroxidase (InterPro:IPR000823), Peroxidases heam-ligand binding site (InterPro:IPR019793), Haem peroxidase, plant/fungal/bacterial (InterPro:IPR002016), Peroxidase, active site (InterPro:IPR019794); BEST Arabidopsis thaliana protein match is: Peroxidase superfamily protein (TAIR:AT5G66390.1); Has 4558 Blast hits to 4529 proteins in 286 species: Archae - 0; Bacteria - 0; Metazoa - 2; Fungi - 246; Plants - 4242; Viruses - 0; Other Eukaryotes - 68 (source: NCBI BLink).	MGLQRILLQGCDASILLDSTLANPSEKENLANGQ SLRGFEVIDEAKAELEFNCPGTVSCADILAFAARD SVLELGGINYQVPAGRRDGKVSIKDEPTQNLPSP NSTVEQLQQMFATKGLNLQEMVALSGAHSIGRV TRCSVFSHRLHTFNATHVQDPSMDLKFAGEMKN NCPSTTSGGEDEVKVPLDGFTPSRLDNRYYRNL ENGRGLLSSDQALMDSPLTAKMVRRFAKYGRV WMPEFVNAMTRMGSIEVLTGSQGEIRRNCRFVN L

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
634	sp_039820_kppp	#N/A	#N/A	-0.8667339	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G11700.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MPPLFSRWYYVCGCLSTGLLFLTFSGVCGWYTE RGNDLVFSLGNSSALIPSDSSEDRIHGFGETSEV LSLSVDLVSCEDLEGVGSFNTTCLLNSNVHLNSSI YAFGIGNLEILPHVLIVCPIEGCSISFNLFGNVKVG PNAAIIAGSVIISSHSLTIGHNSSVNTTSLGGKPPP QTSGTPFGYEGAGGGHGGRGASCLNEQLDAW GGDVYSWSNVSKPWSYGSKGSGRSAQNPFGG NGGGRVKLIVKDELYLNGSVTAEGGDGGVSGGG GSGGSINITAFKLKGLGTISAAGGRGWGGGGGG RISLHCYSVQDVKVTVHGGLSIGCPENAGAAGTY FDATLLSLRVSNDNLSTETETPLLDFSTTPLWSK VFIENSAKVLVPLLWTFRVQLRGQISVLSRSRINFG LMEHPISEFELVAEELLMSDSQIQVYGALRVNVK MLLMWNSEIQIDGGEAVVTTSVLEVRNLAVLREH SVISSNTNLALYGQGLLKLTGDGDCIKSQRLLLSQ FYNITLGRGSLLQAPLDDDASKTMVTKSLCENQT CPMDLINPPDDCHINSTTSFSLQICRVEDILVDGH VKGSIIHFHRARTIVVSADGMITASELGCRKGNG MGRFLKGAGSGAGHGGKGGSGMFNGTISEGGS RYGNANLPCELGSGSEGPNESYGHVAGGGMIVI GATQWPLFRLDLYGSLRADGQSHSQKIKNRDGT LIGGLGGGSGGTILLFLQKLALAEKSSLSAAGGK GGPLGGGGGGGGGRIHFHWSNINTGDEYIPIAVV NGTIYSGGGAGDSSGYYGEEGTITGKECPKGLY GTFCSECPIGTFKDVDGSDEQFCAPCPQELLPH RAEFIYVRGGVNEASCPFKCISDKYRMPHCFTPL EELMYPVGGPWTFSLLLLLGCFVILGLLLNLLRIKM VGSSSVCHVTYSFEQHNFDYQYPYLLSLSEVKG PNKAEDTQSHVYRMYFAGPNTFREPWHLPYSPP DEIVEIVYEDAFNRFIDEINSIAAYDWWEGSVHSIL LVLAYPCAWSWKQWRRRIKIQRLQEYYKSEYDH SCLRSCRARALYKGMKVGGTPDLMVAYIDFFLG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
635	sp_039840_utkd	-0.85889	-0.852133	-0.9k218	LOW-LEVEL BETA-AMYLASE 1 (LBA1); FUNCTIONS IN: in 6 functions; INVOLVED IN: sugar mediated signaling pathway, RNA interference, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, response to sucrose stimulus; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Restriction endonuclease, type I, R subunit/Type III, Res subunit (InterPro:IPR006935), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021), RNA helicase UPF1, UPF2-interacting domain (InterPro:IPR018999); BEST Arabidopsis thaliana protein match is: DNA-binding protein, putative (TAIR:AT2G03270.1); Has 9640 Blast hits to 8330 proteins in 1330 species: Archae - 182; Bacteria - 2840; Metazoa - 3001; Fungi - 1369; Plants - 807; Viruses - 148; Other Eukaryotes - 1293 (source: NCBI BLink).	WDSNANN-YETASQPDTGNDDAYTFLEFNTQAG GDDDDVNFAYPQFNELSQPIRASASPSSSSSA WPTPSDSIVAPEIKESHQPPESHSSSSGGPTKG RGGGGGGGQGVASVVDSLAAGMSGLNFEETAG DEDGFVDQFGKGVGDFSEHACRYCGVQNPACV VRCNVPSCRKWFCNSRGNTSGSHIVNHLVRAKH KEVCLHKDSPLGETILECYNCGCRNVFLLGFISAK TESVVVLLCREPCLSVNALKDMNWDLSQWCPLI DDRCFLPWLVKVPSEQEQLRARQISAQQINKIEE LWKSNPDASLEDLEKPGVDDEPQPVVLKYEDAF CYQNVFAPLIKLEADYDKMMKESQSKDNLTVRW DIGLNKKRVAYFVFPKEDNELRLVPGDELRLRYS GDAAHPAWQSVGHVIKLTAQEEVALELRASQGV PVDITHGFSVDFVWKSTSFDRMQGAMKTFAVDE TSVSGYIYHHLLGHEVEMQVVRNTLPRFGAPGL PELNASQVFAVKSVLQKPISLIQGPPGTGKTVTS AAIVYHMAKQGQGQVLVCAPSNVAVDQLAEKISA TGLKVVRLCAKSREAVSSPVEHLTLHYQVRHLDT SEKSELYKLQQLKDEQGELSSSDEKKYKALKRAT EREIAQSADVICCTCVGAGDPRLSNFRFRQVLID ESTQATEPECLIPLVLGAKQVVLVGDHCQLGPVI MCKKAARAGLAQSLFERLVLLGVKPIRLQVQYRM HPALSEFPSNSFYEGTLQNGVTVNERQSTGIDFP WPVPNRPMFFYVQMGQEEISASGTSYLNRTEAA NVEKIVTSFLRSGVVPSQIGVITPYEGQRAYIVNY MSRNGALRQQLYKEIEVASVDSFQGREKDYIILS CVRSNEHQGIGFLNDPRRLNVALTRARYGIVILG NPKVLSKQPLWNGLLTHYKEHECLVEGPLSNLK QSMVQFQKPKKIYNDRRLFYGAAAGVVPSDNVT PGSTTGASGDRRGSRGRGYMPPGAPNGTHKPA VHPSGFPMTRVPIPPYHGAPPSQPYAIPSRGAVH GPVGAVPQVPQPGNRGFGAGRGSSAPIGSHLP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
636	sp_040130_pjur	#N/A	0.852537	#N/A	RNA polymerase II large subunit (NRPB1); FUNCTIONS IN: DNA-directed RNA polymerase activity, DNA binding; INVOLVED IN: transcription, transcription from RNA polymerase II promoter; LOCATED IN: nucleus, chloroplast, DNA-directed RNA polymerase II, core complex, vacuole; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: RNA polymerase, N-terminal (InterPro:IPR006592), RNA polymerase, alpha subunit (InterPro:IPR000722), RNA polymerase II, heptapeptide repeat, eukaryotic (InterPro:IPR000684), RNA polymerase Rpb1, domain 7 (InterPro:IPR007073), RNA polymerase Rpb1, domain 3 (InterPro:IPR007066), RNA polymerase Rpb1, domain 1 (InterPro:IPR007080), RNA polymerase Rpb1, domain 4 (InterPro:IPR007083), RNA polymerase Rpb1, domain 5 (InterPro:IPR007081), RNA polymerase Rpb1, domain 6 (InterPro:IPR007075); BEST Arabidopsis thaliana protein match is: nuclear RNA polymerase C1 (TAIR:AT5G60040.1); Has 181834 Blast hits to 82224 proteins in 9254 species: Archae - 731; Bacteria - 33255; Metazoa - 56600; Fungi - 34284; Plants - 19037; Viruses - 3576; Other Eukaryotes - 34351 (source: NCBI BLink).	MDTRFFTSPAEVSKVPLDGFGLSPDETRQMSVV QVEFSETTERGKPKPGGLSDMRLGTIDRKLKCD TCTANMAECPGHFGHLELAKPMFHIGFLKTVLSI MRCVCFNCSKILADEDDHKFKQALKIRNPKNRLK KILEACKNKRKCEGGDSIDVQDQDDEEPAKKSR GGCGAQQPTFTIEGMKIIAEYKMTKKKNDEQEQL PEPAERKQQLSAEKVLSVLKRVSDEDCQLMGFN PRYARPDWMILQVLPIPPPPVRPSVMMDTSSRS EDDLTHQLAMIIRHNDNLRRQERNGAPAHIISEFA QLLQFHIATYFDNELPGQPRATQRSGRPIKSICTR LKAKEGRIRGNLMGKRVDFSARTVITPDPNINIDE LGVPWSIALNLTYPETVTPYNLERLKELVDYGPH PPPGKTGAKYIIRDDGQRLDLRYLKKSSDHHLEL GYKVERHLNDGDFVLFNRQPSLHKMSIMGHRIKI MPYSTFRLNLSVTSPYNADFDGDEMNMHVPQSF ETRAEVLELMMVPKCIVSPQANKPVMGIVQDTLL GCRKVTKRDTFIEKDVFMNILMWWEDFDGKVPQ PTILKPRPIWTGKQVFNLIIPKQINLVRFSAWHQE TETGFLTPGDTRVLIERGELLSGTLCKKTLGSSS GSLIHVIWEEVGPDAARKFLGHTQWLVNYWLLQ NAFSIGIGDTIADASTMETINVTISQAKNEVKNLIR AAQEKQLEPEPGRTMMESFENRVNQVLNKARD DAGSSAQRSLDETNNLKAMVTAGSKGSFINISQM TACVGQQNVEGKRIPFGFVGRTLPHFTKDDYGP ESRGFVENSYLRGLTPQEFFFHAMGGREGLIDTA VKTSETGYIQRRLVKAMEDIMVKYDGTVRNSLGD VIQFLYGEDGMDAVWIESQKLDSLKMKKSEFDAT YRYEIDAENWEPDYMLKEHVDDLKSIREFRNVFD AEVQKLEVDRCQLATDIATTGDNSWPLPVNLKRL LWNAQKTFKVDMREKSGMHPMEIVEAVDKLQER LRVVPGDDFLSKEAQKNATLFFNILLRSTLSSKRV LKEFRLNREAFDWVIGEIESRFLQSLVAPGEMIG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
637	sp_040150_kxyz	#N/A	#N/A	-0.906885	P-loop containing nucleoside triphosphate hydrolases superfamily protein; BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT2G19120.1); Has 5876 Blast hits to 5615 proteins in 1104 species: Archae - 176; Bacteria - 1754; Metazoa - 1255; Fungi - 973; Plants - 640; Viruses - 179; Other Eukaryotes - 899 (source: NCBI BLink).	MGSRGRLLFDLNETPVEVDEEIDFPSGVLPQKVL PSLNSCTSDLLVSSVAPTSIVNNNAFSHASSVSV FQPFVRIKESGLDRIVDQQRARDMDCNTSLPEVD ENSDRKEVMKVNSGNIYAQVIEKEEGEWSDGEC SEPHVSQRMEDNSSASGNKDLNEQGIPDVAANN SSSGTSLKKNSRDSSPADSRKIKHDYASAKPDTF PNDITGGTSQNMNGNAERNTSVDFQGDSESVSK QKEVKGAEAIHALKSANSLGKHKLDEHKEAMLGK KRSRKTMFLNIEDAKQAGAVKSSTPKRQTVPTPS STRTVKGSRNALLSAERNGEKQSEPAIRDQNQV ELSCNEESNPMEPIDRKSESNSDVNAGLLARPR RLNSCNDLSTEGNTRHNLWKQPTDPRQFKDPQ TQPRKLASSSHGSGLVDVKQLNKKNPPKRQLVT NHQYQDTSVERLLREVTSEKFWHNPEECELQCV PGRFESVEDYIRVFEPLLFEECRAQLYSTWEELT ETSARDAHVMVRVKSVERRERGWYDVVVLPAN ENKWTFKEGDVAVLSCPRPGSVRSRRNSTSSAE EYEEPEGSARTVGTVRRHTPLDVRDPQGAILHFF VGGSYDPNSAADSDHVVNKLQPRSIWFLSVLGS LATTQREYVALHAFRRLNVQMQNAILKPSAEQFP KYQEQTPAIPECFTPNFVDHLHRTFNDPQLSAIQ WAAMHTAAGTSNGVTKRQDPWPFTLVQGPPGT GKTHTVWGMLNVIHLVQYQHYYTALLKKLAPESY KQNNERNADTGSAGSIDEVLRSMDENLFRTLPKL CPKPRMLVCAPSNAATDELLARVLDRGFIDGEMK VYRPDVARVGVDTQSRAAQAVSVERRTEQLLVK TRDEIFGWLQQLRGREAHLSRQMATLQVQLNSA AAAGRSQGSVGVDPDVLMARDQNRDALLQTLA AVVEERDKVLVEMSRLLILEGKFRSASSFNLEEA RASLETSFANEAEVVFTTLSSSGRKLFSRLTHGF DMVVIDEAAQASEVGILPPLALGAARCVLVGDPQ QLPATVISKAAGTLLYSRSLFERFQQAGCPTMLL
638	sp_040530_xuja	-0.954309	-0.924313	-0.878287	ECERIFERUM 7 (CER7); CONTAINS InterPro DOMAIN/s: Exoribonuclease, phosphorolytic domain 2 (InterPro:IPR015847), Exoribonuclease, phosphorolytic domain 1 (InterPro:IPR001247), Ribosomal protein S5 domain 2-type fold (InterPro:IPR020568); BEST Arabidopsis thaliana protein match is: ribonuclease PH45A (TAIR:AT3G12990.3).	MEQRLANSWNMTVNEKNFIETALLSDLRVDGRR PFDYRRLTIKFGREDGSSEVQLGQTHVMGFVTA QLVQPYKDRPNEGSLAIYTEFSPMADSSFETGRP GESAVELGRIIDRGLRESRAIDTESLCVLAGKLVW SIRIDLHILDNGGNLVDAANIAALAALLTFRRPECS LGGEDGQELTVHPPEVREPLPLIVHHLPIAVTFAF FTADTVVVIDPSHSEEAIMKGRMTVTLNTNGDICA IQKGGGVGVMHNVVMQCLRVASVKSADITTKIKE AVESYNTEKAYRKVKRHPSSAGGQTER

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
639	sp_040550_soaa	#N/A	#N/A	-0.3631141	unknown protein; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF814 (InterPro:IPR008532); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MVFYFKARPEAGDYTIFMGLDKFENEELIKYGFP EDIWFHVDKMSSAHVYVRLHRGQTMEAISEGLL EDCAQLVKANSIQGNKVNNIDVVYTPWANLKKTA SMDVGQVGFHNSKMVRTVRVEKRINEVVNRLNR TKVERTPDFKAEREAYAAAEKAERKAQQREKKR REEMDRLEKERVSEMRSYKNLMVSDKMTSNKQI ASAEKSLQELEEDFM
640	sp_041120_iasz	-0.8624206	-0.947361	-0.8722373	uncharacterized protein	MPGVTADSTVVEGDLDSSPQRTSSAESNFRELD DAYLQNQARIWLGEVLQIRFDEQIDISDLLADGEL LFEVAEVISRTLPTLSADESHIKAHERKTLASNKS TRRYMPYYNVDSFLRMCKNLGIDGIDLFTPSDVV ERRNTRKVCMCLRSLSKKKSRLRDLNVPDFDSVT CTLAMPTDMVECIRKSLEKSHSSSSLSVEQEAEK VSRVRVWFRMSNSASAVSNGYSDSCDEAESKF VVAESYGSSDDAYDAESPDISGSQVVREYNSEL QLSTRSDFENEDDDLCEYLSPSNAESVGSPCSQ YYSDEDLSVPSSNGLGSRISGKVPELNIETNHEY QISDVVHVPIDWEDIHLDVGNNKGVNANDGEVD CVSHDNEDVDTTCSHGFYLIGNDMCNEQLGTSN NPSKVDCDNISELGEPPECVLTDDFHVSDAPHLS HEDAPSENNATKHDPLVCEVDLGTGEHYSHGKT KDTVLYGELLIQSSCVSKPSESPDLLYDSENQAS NACTKVINIGVNSCCDRVNNENVTSEVVAGELRF HEEQSPCSVDYDGNTFNCSSLVDVKSNDYDDHA TFVDPIRTSIPSLFERDTVICTDVHCSPIENDKESL ESKMMPDNKEQHSDNGCMEGNASNESDMPQM DLNSSPCMVEAKNITGCSKGCHDTMEGSEGHNA LVESVDHKPLDNTILVSSFATAVCDPHSEAQHPKI LEDTDHKHLQCVKEAEKGESKFVKQEGTNKEVP VHKFQRKLLKSIVKGTTLVGVAVFLLHLRKNRRE NTSGAKKRPAQIVIQGTTKPLQKGQMGGGTSSV YPAEKLHLKL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
641	sp_041560_zjad	#N/A	-0.873306	-0.86334	Chaperone DnaJ-domain superfamily protein; FUNCTIONS IN: heat shock protein binding; INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Heat shock protein DnaJ, N-terminal (InterPro:IPR001623); Has 3526 Blast hits to 2426 proteins in 300 species: Archae - 4; Bacteria - 245; Metazoa - 1133; Fungi - 271; Plants - 159; Viruses - 2; Other Eukaryotes - 1712 (source: NCBI BLink).	MGAPKASSHDVMPEKNFDERILKKCTSRKRDNIV LIEVDSDDFGDVIILDVPDSFQKKFRKRAKLVINID DEDGGGSSDGVCHTNSDATTSDQCFPTEAADP STTSYDVNGEDTPVKLSKSKRTYPGKSIPRNHCG ISLDSESCSSESDSSDCVLMEGSSGKLREHWEK ASFRRRQDASCAPHDFENEAGTSVPSHPVHDYA EDAEPPVQQSNVRKCSVQQDGDNVFAGYPEKF SEESKASDQDYDPMEIDGHVNSSSDEKERGGTL SPESPVHHNTEEPAGDFRDSTGKLAEELKFPRP DYGDMTKNEVCSVYSTSITKGPSEDVESDHATLI DRQQGVEIVCMLPDHELGCGNQVVSSSDQINIID RQKIKETEEYKHAMEEELAARQQQLLIQAEEARR LRNRKKAENMRIQMNERRQKERVEEIRQTRRRD EENLNLKEKHRSEIKMELRRTEAICSDMASLLRCL GIHVGGGQYPPPNEVQAAYKRACLKFHPDRMST TCLRQQVEAEEKFKLISNMKDKFLQNR
642	sp_042010_qucd	-0.869532	#N/A	#N/A	SUPPRESSOR OF GAMMA RADIATION 1 (SOG1); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: multicellular organismal development, response to gamma radiation, DNA damage checkpoint, regulation of meiosis; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: No apical meristem (NAM) protein (InterPro:IPR003441); BEST Arabidopsis thaliana protein match is: NAC domain containing protein 85 (TAIR:AT5G14490.1); Has 2166 Blast hits to 2160 proteins in 77 species: Archae - 0; Bacteria - 0; Metazoa - 22; Fungi - 5; Plants - 2134; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	RSWLVDGNRFATKIKCASGSCDLGNIKWKSNPT RSCPNCHHTIDNNDVGQEWPGLPRGVKFDPTD QEITQHLLAKSGIGDEAPHPFIDEFIPTVDEDDGIC YTHPENLPGVKRDGSASHFFHRAIKAYNTGTRKR RKIHGNNLGDVRWHKTGRTKPVLLNGEQKGCKK IMVLYVAVKGGKPEKTNWVMHQYHLGIEEDEKD GSHVISKVFYQQQQLVKQSEKTTGYISDVAEDSV VKVDPVTPKSVTPDPPRNEMRLHTKNDISSVNC NSGAQDREMGFADEVQQEITISSGVNDSIKDDCI DITSQDPVLEHGDNTVEEEKWWDSESQYLLNSQ QLVEGLSLCDELLQSQSPNRNGSTNTDDVAKKR YVLSEYARLGPDDLKKDLEDCQNLVLDPANIDLD TPPDFRLSQLEFGSQDSYSAWGLGKTTD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
643	sp_042030_fpyg	#N/A	#N/A	-0.8493776	Transducin/WD40 repeat-like superfamily protein; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat, conserved site (InterPro:IPR019775), WD40 repeat (InterPro:IPR001680), G-protein beta WD-40 repeat, region (InterPro:IPR020472), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT5G66240.1); Has 22372 Blast hits to 13576 proteins in 557 species: Archae - 62; Bacteria - 5277; Metazoa - 7900; Fungi - 4335; Plants - 2143; Viruses - 0; Other Eukaryotes - 2655 (source: NCBI BLink).	MAITSLTELDDDVVRSMAIGAVFSDFGGQINSIDF HRTSDLLVTASADDSVRLYDIVNAKLLKTTYHKKH GADRICFTHHPSAVVCSSKYNLDSTGESLRYLSM YDNRCLRYFKGHKERIVSLCMSPINDSFMSGSLD HSVRIWDLRVNACQGILRLRGRPAVAYDQQGLV FAVAMEGGAVKLFDSRSYDKGPFDTFLVGGDNA EVCDIKFSNDGKSMLLTTTNNVYVLDAYAGEKR CGFSLEPSQNTTIEATFTPDGQYVVAGSGDGSM HAWNINMRSEVARWSSHIGVPSCLKWAPRRVM LVAASSVLTFWIPNHGKITGESNANDADVATKLE VLPH
644	sp_042100_dsea	-0.930885	-0.9111061	-0.9113988	uncharacterized protein	FLFSLSSSSHLIFSLFYSSFHHKRSHTSAAVSIASP PPLTSLPLPIDLVADISPLSNSSVISLPPQAQAQR RSPPSYYLTLSWDLRKHEIKSGRLWIMLESHNLN HPWLILLNPLRFKLTCVVN
645	sp_042830_ozyn	-0.9313223	-0.9611591	-0.9327497	uncharacterized protein	MVELTPNLNFRDTKEEEDLQVDNRSQPSPIRREK ASPTVKFEFEEQIAINQSPQIVSDPSLDGYAVHKL GVIQFL
646	sp_042930_kryi	#N/A	#N/A	-0.866811	Protein of unknown function (DUF1423); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1423, plant (InterPro:IPR004082); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G14840.1); Has 192 Blast hits to 189 proteins in 31 species: Archae - 0; Bacteria - 0; Metazoa - 2; Fungi - 10; Plants - 173; Viruses - 0; Other Eukaryotes - 7 (source: NCBI BLink).	MTCENEDASGHGGELMEVDIADASNGCMTEIRE NGLLLWPVLPQESGEGLPYAPVNWPKPGDWG WRVGKRVAASGYYLDRYLYLPLSYQRSRRKRDG FASKLSVEQYVRARFPGTNVYAFFSSFSWKIPSK QAYLLKDEHGSFRISSEEEMAAQSPPYMLPGGLG CKAGNIFCASLLEDQMSPSATLSCDICCSEPSFC RECCCILCCKTIDLGLGGYSFIRCEATVHDDLICG HAAHINCALRSYMAGTVGGSVGLDTEYYCRRCD TRTDLTKHVTRLLKTCESIESVDEIDKMLSVGACI LRGSQKKYSQKLLNRIESILHKLKFGNDVKDVWK VEVEILPVPTPTPTPTGQLDYANSACTDVGEDDL LYQNGYLKLEEEINQVLQELKSSQEYEFSIAHDKL HAQKDNLFSLYQQVYKDKLELLLRPSPSKDRDALL RSIAAREEYIKHELSKLRDMEKVANGFGKTPKKLL KEHFDLDLKD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
647	sp_043190_gcix	-0.947319	-0.944454	-0.933866	nine-cis-epoxycarotenoid dioxygenase 5 (NCED5); CONTAINS InterPro DOMAIN/s: Carotenoid oxygenase (InterPro:IPR004294); BEST Arabidopsis thaliana protein match is: nine-cis-epoxycarotenoid dioxygenase 2 (TAIR:AT4G18350.1); Has 2882 Blast hits to 2845 proteins in 494 species: Archae - 16; Bacteria - 802; Metazoa - 232; Fungi - 202; Plants - 893; Viruses - 0; Other Eukaryotes - 737 (source: NCBI BLink).	MMNPALTNPCTKSDWAQLAFNSSPKKAMICTRK LSKENIGYNYYTSSNNNIQCATASPSILHFPNTTH DNNVKNPPTKPQQLKKWNIFQKAVGAAIDMVET ALVSLERSQPLPKTSDPKVQIAGNFAPVPEMPVK QSLHVTGEIPGCIDGVYVRNGANPLFEPTAGHHL FDGDGMVHAVTIHDGVASYACRFTETSRLAQER DLGRAVFPKAIGELHGHSGIARLMLFYSRGLFGLI DNEHGMGVANAGLVYFNDRLLAMSEDDIPYQVR VTPAGDLKTVGRFDFKGQLTSTMIAHPKLDPESK ELFALSYDVLKAPYLKYFKFSQDGEKSADVDIPLN EPTMMHDFAITKKFVVIPDQQVVFKMKEMITGGS PVVYDKNKTSRFGIMRKDAKDASDITWVTSPHTF CFHLWNAWEESETEEIVVIGSCMTPPDSIFNECD ENLQSVLSEIRLNLRTGESTRRPIIDFSQHPHINLE AGMVNRNLLGRKTRFAYLAISEPWPKVSGFAKV DLITGEVKKFMYGSEKYGGEPFFLPHDNSILDES EDDGYIMGFVHDEEECKSELLIVNAQTLQLEAAV KLPSRVPYGFHGTFVRSEDLSHQA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
648	sp_043260_pwje	-0.8837634	-0.8837009	#N/A	SUPPRESSOR OF AUXIN RESISTANCE 3 (SAR3); FUNCTIONS IN: porin activity; INVOLVED IN: mRNA export from nucleus, defense response signaling pathway, resistance gene-dependent, response to auxin stimulus, negative regulation of flower development, developmental process; LOCATED IN: nuclear membrane, nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase S59, nucleoporin (InterPro:IPR007230), Nuclear protein 96 (InterPro:IPR021967); BEST Arabidopsis thaliana protein match is: Nucleoporin autopeptidase (TAIR:AT1G10390.2); Has 653 Blast hits to 526 proteins in 148 species: Archae - 0; Bacteria - 0; Metazoa - 269; Fungi - 224; Plants - 93; Viruses - 0; Other Eukaryotes - 67 (source: NCBI BLink).	MECDTLRDPFSSRTFNFEYSSSLQISSPRKKRKT NRVKYDGFFPILRSSDYYTEPPLKELAELEEIYPG YCSKIKDFTVGRVGFGSVKFLGETDVRRLDLREI VKFNRHEVVYEDESSKPDVGEGLNKPAVVTLFL KFETLSFDRSKFCKVVDRLRASTDRQGAEFISYA PLNGEWKFSVRHFSRFGLSMDDEDDIVMDDADV NPCEDPVEMNGGVGFGDGEDDDICSDGFDMGE NSAELSHSLPAHLGLDPARMKEMKMLMFSGEDD EDENYDDILSLEKQGAYAEKFRLYNSGLRVNYKP SPPPPRKTPVPLLEYKSNSLDVNAPGTLLLSQQH KGSPLRVVKPDGFKLNYKQETPVSGYHSRNIVDA GLFMGRSFRVGWGPNGLLVHSGVPVSSNDSQG ILSSVITLEKVAIDKVTRDESNEVKKDAVDSLFKSP LDLHKSLDHRILEVEVGSLKLSLQEVVCDRSALS DICRRYIEVVERQLEVPGMLPPVRMSLMHQVIVW ELIKVLFSERETSGRTVFGVEDHEEDMMLDGKAI RSEIDPESLPLMRRAEFSYWLQESVYHRVQEDL SCLSESDDLEQIFLFLTGKQLDAAVELSASRGDV RLACLLGEAGGSMVNRSDVARQLELWRTNGMD FNFIEKERIRLFELLSGNIHSALAGKSIDWKRFLGL LMWYELPPGTSLPAVFHTYQHLLIEGKAPYPVPV YIDEGPVEETVKVDTEGRYDLAYYLMLLHASEDG KFDGLKSMFSAFASTNDPLDYHMIWHQRGILEAI GAFSSSDLYILDMAYVSQLLCLGKCHWAIYVVLH MPYCDDFPYLHSTVIREILFQYCDTWSTQEEQRK FIEELGIPSQWLHEAMAVYFQYYGDLAKALEHYIR CFCWQKAHSIFITSVAHSLFLSAEDSEIWRLAMS LEDHKSEIEDWDLGAGIYISFYLLRSSLQCDEDDI NKLDTLESKEAACREFFDCLNKSLEVLGSRLPTD GRVVYSKMSEEICSLLVSDSGADSTRETQLSCFD TIAAAPIPEDLRSCHLQEAVLLFTTFLSETTT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
649	sp_043420_orwr	-0.878124	#N/A	#N/A	embryo defective 1075 (emb1075); FUNCTIONS IN: pyridoxal phosphate binding, carboxy-lyase activity, catalytic activity; INVOLVED IN: cellular amino acid metabolic process, embryo development ending in seed dormancy; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Pyridoxal phosphate-dependent transferase, major domain (InterPro:IPR015424), Pyridoxal phosphate-dependent decarboxylase (InterPro:IPR002129), Pyridoxal-phosphate binding site (InterPro:IPR021115), Pyridoxal phosphate-dependent transferase, major region, subdomain 1 (InterPro:IPR015421); BEST Arabidopsis thaliana protein match is: glutamate decarboxylase 2 (TAIR:AT1G65960.1); Has 3259 Blast hits to 3251 proteins in 995 species: Archae - 187; Bacteria - 1878; Metazoa - 518; Fungi - 211; Plants - 249; Viruses - 11; Other Eukaryotes - 205 (source: NCBI BLink).	MAGLAAFPQELSQDVVDLLNRTDLDPTAVVFDPI PPVVVTSNGENGVVKMGREIVMGRNIHSSCLDIT EPEANDDVTGDKEAYMAGILARYRQSLLERTKYH LGYPYNLDFDYGALSQLQHFSINNLGDPFIESNY GVHSRQFEVGVLDWFARLWEIEKDQYWGYITNC GTEGNLHGILVGREVFPDGILYASKESHYSIFKAA RMYRMDCVKIDTLSSGEVDCADFRTKLLQNKDK PAIINVNIGTTVKGAVDDLDLVIQTLEECGFPQNR FYIHCDGALFGLMLPFVTRAPQVTFKKPIGSVSVS GHKFVGCPMPCGVQITRMDHINTLSSNVEYLASR DATIMGSFNGHAPIFMWYTLNRKGYRGFQKEVQ KCLRNAHYLKDRLRDVGVSAMLNELSSTVVFER PRDEEFIRRWQLACQGEIAHVVVMPSVTIEKLDT FLDELVEKRLTWFKDGGNQPPCLASEMGKDNCA CPLH
650	sp_043640_gfmq	-0.8804[illegible]	#N/A	#N/A	Transcription factor TFIIE, alpha subunit; FUNCTIONS IN: RNA polymerase II transcription factor activity, transcription initiation factor activity; INVOLVED IN: transcription initiation from RNA polymerase II promoter; LOCATED IN: endomembrane system, transcription factor TFIIE complex; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Transcription factor TFIIE, alpha subunit (InterPro:IPR002853), Transcription factor TFE/TFIIEalpha, HTH domain (InterPro:IPR017919); BEST Arabidopsis thaliana protein match is: Transcription factor TFIIE, alpha subunit (TAIR:AT4G20340.1); Has 759 Blast hits to 706 proteins in 191 species: Archae - 5; Bacteria - 10; Metazoa - 313; Fungi - 190; Plants - 108; Viruses - 11; Other Eukaryotes - 122 (source: NCBI BLink).	MSNSLSIEPFNRLVKLTARAFYDDISTKGDNQPK TSRSDNRGTAVVILDALTRRQWVREEDLAKDLKL HLKQLRRTLRFFEEEKLVMRDHRKETAKGVRMH NAAVSATANHHYGREGDDKIKLHTHSYCCLDYA QIYDVVRYRMHRMKKKLKDELDSKNTIQEYICPN CNKRYTALDAARLISEDGENFLCESCQGELVAES DKLAGQDVGDADENVRRRRRREKLKEMLQKMEE QLKPLVEQLNRVKDLAVPDFGSLQAWEVKASAA ARAAGGDNPDDPSRSYQGGTPMPFLGDPKIHVD ISGVEDKEDSKSGTGSAPMKVLPPWMIRQGMKL TKEQRGEGEEESKPDVTSEPAGLSDDKKSTVVK DDKSIQDEYFKAYYAAILERQRETDVKGEPDVST EEVTGQNTSLDRQVGLKSKREEQSEDEDVWEE ATPSAGNTGESYKVDLNVQAEPLPEEEEDNDEE DVEWEDG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
651	sp_044010_jqjo	-0.877833	#N/A	#N/A	IND1|iron-sulfur protein required for NADH dehydrogenase)-like (INDL); LOCATED IN: mitochondrion; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: ATPase-like, ParA/MinD (InterPro:IPR019591); BEST Arabidopsis thaliana protein match is: ATP binding (TAIR:AT3G24430.1); Has 13193 Blast hits to 13164 proteins in 2587 species: Archae - 550; Bacteria - 7572; Metazoa - 442; Fungi - 431; Plants - 201; Viruses - 0; Other Eukaryotes - 3997 (source: NCBI BLink).	MKGAFKINTVFNNLRNVGIRGLSAMAGKAKGLQI SGVKDIIAVASGKGGVGKSTTAVNLAVALANRGQ LAVGLLDADVYGPSIPTMMKLHGKPEATEDLKMI PKENYGVKCMSMGLLVEKDAPIVWRGPMVMSA LEKMAKGVEWGKLDILVVDMPPGTGDAQLTISQ RLQLSGAIVSTPQDIALLDARRGAHMFSKVNVPII GIIENMSCFMCPHCSQPSFIFGEGGAKRTATEMS LDYLGDIPIEMDIRTCCDEGVPIVLSSPDSTSAMK YTDVAQKVMKRIEEISHKQFSPEISL
652	sp_044230_meqy	-0.886578	-0.93373	-0.875226	GNARLED (GRL); CONTAINS InterPro DOMAIN/s: Nck-associated protein 1 (InterPro:IPR019137); Has 2387 Blast hits to 741 proteins in 151 species: Archae - 0; Bacteria - 54; Metazoa - 569; Fungi - 142; Plants - 67; Viruses - 0; Other Eukaryotes - 1555 (source: NCBI BLink).	MSEWRRKISWQRGHVDENLQNQMQSSSWTAD MKLKQVPSLLYLFFILDCWHEANRSDLSPKLVFV DQLCEISPHLPRSSRWNSIIRSLYNQYYGNSSSL PLTLRL
653	sp_044830_weky	-0.952723	-0.886438	-0.914431	SBP (S-ribonuclease binding protein) family protein; FUNCTIONS IN: zinc ion binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841), S-ribonuclease binding protein, SBP1, pollen (InterPro:IPR017066); BEST Arabidopsis thaliana protein match is: SBP (S-ribonuclease binding protein) family protein (TAIR:AT1G60610.3); Has 1212 Blast hits to 1212 proteins in 167 species: Archae - 0; Bacteria - 2; Metazoa - 524; Fungi - 0; Plants - 416; Viruses - 105; Other Eukaryotes - 165 (source: NCBI BLink).	MFGGNNSNHAFPMFLDENRFQLNTNAPNHQLQL FGNX

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
654	sp_044870_omfi	#N/A	-0.8503119	#N/A	CRINKLY4 related 1 (CCR1); FUNCTIONS IN: protein serine/threonine kinase activity, receptor activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: TNFR/CD27/30/40/95 cysteine-rich region (InterPro:IPR001368), Protein kinase, ATP binding site (InterPro:IPR017441), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Regulator of chromosome condensation/beta-lactamase-inhibitor protein II (InterPro:IPR009091), Protein kinase, catalytic domain (InterPro:IPR000719); BEST Arabidopsis thaliana protein match is: CRINKLY4 related 2 (TAIR:AT2G39180.1); Has 92283 Blast hits to 90862 proteins in 3897 species: Archae - 88; Bacteria - 8648; Metazoa - 34704; Fungi - 6843; Plants - 29398; Viruses - 212; Other Eukaryotes - 12390 (source: NCBI BLink).	MLSLSSLFLLFLLSLPISAYAFGSMGPISAAFGGG DAFFCAIDASGKQGIICWGMSPEKTSENFTTIPS MAALSGGEGFMCGILANTSQAYCWDSHGTGNE LVPPFYRYNNYSSIASGKNHVCAIRGNYYSDSSN SGSVDCWEIINGSNSNTFTSKQSALFYDQSISSL VFEKIVSGDGFSCGIVKDDEELICWGPSSTNLGIS KISDNFLGLSSSRSSVCGVVKPSGEVKCWGKND SYPDDPSGVRFVALTGGSRHFCGIRADTHAVDC WGGFNSSSIPKGSGFFAIATSDVAACGIREEDLV LDCWFVNESSAPTQYNPPLQLCSPGLCSPRSCD VGEFTFNASALNEPELTNVCIRRDLHLCSSCGTN CSSGYFLSSPCTGNADRVCTECSLCQNSSCWG VCGLAEVELKRRHQRQMQKIVLILGPSIAGFLILI GWCLFPYLYFFKAKDGPTKQFRSCIGKKDPESD ANGDANPPMAVACCPDEAQIFRLSELKDATNGF KEFNELGRGSYGFVYKALLTDGRQVAVKRANAA TIIHTNIREFENELEILCKMRHCNIVNLLGYCTEMG ERLLVYEYMTHGTLHDHLHDGLSPLSWTIRLRIA MQAAKGLEYLHMEANPPVVHRDVKTYNILLDSD WGARISDFGCLNFNERDLDEDMKNDVVNFGVVL LEILSGRKAYDVDYSPPNIVEWAVPLIRRGKAAAL IDRYVALPRNVEPLLKLADIAESAVRESSSDRPTM SAIANSLEQIVKEGLIL
655	sp_045040_dqja	0.9944036	0.8083573	0.8855857	uncharacterized protein	VPPPLTAVVPPAAGRCIPAVPHRRVAPPCIL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
656	sp_045120_ufuo	-0.884377	#N/A	-0.911742	chromatin remodeling 8 (CHR8); FUNCTIONS IN: helicase activity, DNA binding, ATP binding, nucleic acid binding; INVOLVED IN: DNA repair, response to gamma radiation; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021), SNF2-related (InterPro:IPR000330); BEST Arabidopsis thaliana protein match is: chromatin remodeling 24 (TAIR:AT5G63950.1); Has 17230 Blast hits to 14463 proteins in 1653 species: Archae - 99; Bacteria - 4180; Metazoa - 3846; Fungi - 4219; Plants - 1677; Viruses - 118; Other Eukaryotes - 3091 (source: NCBI BLink).	WEEEDAILNSLGVTSANPDDIERDICNKATTNV ESSSQAGRRTTEADDFTDTTEDQDSSSSQVELL KKLRAIEVEIHAVASAVEQTKSGGKNEELDSESH GKKEEDDEGIDKAIPRDSALQHALAADRLRSLKK TRAELRKELSSICQDGKTVGVEEKLLKYLVREES KSKRKLKEVQGPSKESKKKKAVSYQEDDEFDAV LNAASGGFVETERDELVRKGLFTPFHKLKGFERR IQQPETSSLQQNGNDIADRASQSVAKAVRSLTEA AQARPTSKLLDPKDLPRLDAPSFAFNRLSKSFKV SKSLDDGEKSNDAKRKKRRPLPGERWRKRIAKE EIEVEEPDVQDDLITGTQDADNEDGDVTPSVTLE GGLKIPETIFDKLFDYQKVGVQWLWELHCQRAG GIIGDEMGLGKTIQVLAFLGSLHFSKMYKPSIIVCP VTLLRQWKREARKWYRSFHVEILHDSAVDGGRK QKQSESDESENESEGSSESDQDRVPPSKSAKK WDSVIKRVLRSESGLLITTYEQLRILGEKLLDIEW GYAVLDEGHRIRNPNAEVTIVCKQLQTVHRIIMTG APIQNKLSELWSLFDFVFPGKLGVLPVFEAEFAV PIQVGGYANASPLQVSTAYRCAVVLRDLIMPYLL RRMKADVNAHLPKKTEHVLFCSLTPLQRSVYRAF LASSEVEQIFDGNRNSLYGIDVMRKICNHPDLLER EHASQNPDYGNLERSGKMKVVAQVLKMWKEQG HRVLLFAQTQQMLDILENFLISNEYSYRRMDGLT AVKQRMALIDEFNNSNDVFIFILTTKVGGLGTNLT GANRVIIFDPDWNPSTDMQARERAWRVGQTKD VTIFRLITRGTIEEKVYHRQIYKHFLTNKILKNPQQ RRFFKARDLKDLFVLTDEEHAGSTETSNIFGQLS EQVNVISKEKEDQQNNNCAGATAACADDMELDR ENSSQQELSGNKEKEEAENGGEADEETNFLKSL FDTHGIHSAINHDAIMDAHDEEKIRLEEQASLVAQ RAAEALRQSRMLRSKESISVPTWTGKSGAAGAP SSVRRTFGSTVNSQLVRNSRSSEEGAPKNGTTN
657	sp_045280_zpgr	-0.8869874	#N/A	-0.8869879	RNA polymerase III RPC4; FUNCTIONS IN: DNA-directed RNA polymerase activity, DNA binding; INVOLVED IN: transcription from RNA polymerase III promoter; LOCATED IN: DNA-directed RNA polymerase III complex; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: RNA polymerase III Rpc4 (InterPro:IPR007811); BEST Arabidopsis thaliana protein match is: RNA polymerase III RPC4 (TAIR:AT4G25180.1); Has 234 Blast hits to 234 proteins in 100 species: Archae - 0; Bacteria - 0; Metazoa - 103; Fungi - 41; Plants - 59; Viruses - 0; Other Eukaryotes - 31 (source: NCBI BLink).	MDPASFDQKPRVPRKVRFQPKAPVRKPSKAAVA KTEVADDAEAAQKKELMRRFQDGLEAKPKFEKK TERVQRVAFGYGSSIPGTQFSASNGNNSWRQD KAHSNLRENKEYKEPWNYYSYYPVTLPLRRPYS GNPDLNDEEFGQLSTFDESTSNPAMELGLMGE APEPRMMFLQLPTTIPMTKRNSAAAGQESTSSS KPPSGSNPPVEKAAGLKELQAGVMGKMLVYRS GAVKLKLGDTLYDVSLGSDCAFAQDVVAINAVDK HFSVVGELNKRAILTPDVDSILDSFADLG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
658	sp_046480_fmth	-0.899727	#N/A	-0.850181	ENTH/ANTH/VHS superfamily protein; FUNCTIONS IN: phospholipid binding, clathrin binding, binding, phosphatidylinositol binding; INVOLVED IN: N-terminal protein myristoylation, clathrin coat assembly; LOCATED IN: membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Epsin-like, N-terminal (InterPro:IPR013809), ANTH (InterPro:IPR011417), ENTH/VHS (InterPro:IPR008942), Clathrin adaptor, phosphoinositide-binding, GAT-like (InterPro:IPR014712); BEST Arabidopsis thaliana protein match is: ENTH/ANTH/VHS superfamily protein (TAIR:AT1G14910.1); Has 1020 Blast hits to 994 proteins in 190 species: Archae - 0; Bacteria - 14; Metazoa - 316; Fungi - 202; Plants - 440; Viruses - 0; Other Eukaryotes - 48 (source: NCBI BLink).	MATLTWRKAYGALKDQTKVGLATINSDYKDLDV AIVKATNHVECPPKERHLRKILLATSAIRPRADVA YCLHALARRLGKTRNWTVALKTLIVVHRLLREGD PTFREELLNFSQRGRVLQLSNFKDDSSPVAWDC SGWVRTYALFLEERLECFRVLKYDIEAERLPKPA EGQEKGYSRTRDLGSEELLEHLPALQQLLYRLIG CRPEGAALGNYVIQYALALVLKESFKIYCAINDGII NLVDKFFEMPRHEAVKALEIYRRAGQQAASLSEF YEVCKGLELARNFQFPVLREPPQSFLATMDEYIR EAPKMVDMPAQQLEFPERLLLTYKPEENSLSED EKSNVEEPPPMASDDLTSSAEAALATTRALSPAT NMDSGDLLGLNDVTPNATAIEEGNALALAIVSSD AASTSNSGTVSSSFDPSGWELALVTAPSTDISTH TERQLGGGLDSLTLNSLYDEAAYRANQQPVYGA PAAPNPFEVHDPFGMSHNIAPPTAVQMNSMTQQ SAPNPFGPYQAPFHAQSQPPVGLGAPNPFGDM GFGPYAVTPVNHPQHAANPFGNTGLL
659	sp_046590_kwap	#N/A	-0.830169	#N/A	CRUCIFERINA (CRA1); FUNCTIONS IN: nutrient reservoir activity; INVOLVED IN: response to abscisic acid stimulus; LOCATED IN: endomembrane system; EXPRESSED IN: guard cell, seed; EXPRESSED DURING: seedling growth, seed development stages; CONTAINS InterPro DOMAIN/s: 11-S seed storage protein, conserved site (InterPro:IPR022379), Cupin, RmlC-type (InterPro:IPR011051), Cupin 1 (InterPro:IPR006045), RmlC-like jelly roll fold (InterPro:IPR014710), 11-S seed storage protein, plant (InterPro:IPR006044); BEST Arabidopsis thaliana protein match is: cruciferin 2 (TAIR:AT1G03880.1); Has 1047 Blast hits to 1009 proteins in 168 species: Archae - 0; Bacteria - 115; Metazoa - 2; Fungi - 0; Plants - 930; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MASTTTNALLFLACLCLVALFPPTLAQTPLQQPQ QPLRVRGQQLQHECDIQQLTAAEPTRRVRAEAG VIEVWESNTEEFRCTGVAAVRYVIEPKGLLLPSY TNAPYVTYVMQGKGIQGVILPGCPETFESPVRG GSEGSKGVERDQYQRVLRVHEGDVIGSLAGGV QWTYNDGDTPIVSITLLDLSNPQNQLDLNFRSFY LAGNPQGQGGQEQGPKEVVGKNIFNGFDDELLA DAFNVDTETIRSIKGENDERGSIIRVERDLELLSPE WDDMEEERMPRLNGLEQVLCTISFKQNIDQPSR ADVYTEHGGRINTLNSHKLPLLQHLQLSVERGVL YKNAMMTPHWNINAHSIIYITRGTGRIQVAFENGIL VFDGRVQEGQLLVVPQNFVVLKKAEQEGLEWVS FKTNDNAMISPLAGKLSAIRGMPEEVLMNSYGLS RDEVRRLKYGREQLSLFSPRLD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
660	sp_047050_qjrz	-0.890644	#N/A	-0.903181	CONTAINS InterPro DOMAIN/s: WRC (InterPro:IPR014977); Has 219 Blast hits to 219 proteins in 19 species: Archae - 0; Bacteria - 0; Metazoa - 3; Fungi - 0; Plants - 215; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	MRIRNRKVPLPLSLLSPVPLSDPHLYSRSPVMPV QDVATYSPEPPPSPNDLTTPTSWPSSDLHQHRG LDNNTNVGATRFEDGEQVQISSPAPILHQLSATT SAAGASENNNGRLVFLSNNKDRVGRWCDGDKAI PLKKRRGFGFNNTSFAENSSTRITTDEDEDEDED EDEDEDEDEDEDEDDDTATTATTEMIIMKWNNIN DDEEEKQAVKDDKDAVVAPPAPAPAPAPAPAPA PAPAPATTTVKKGNAKRGNTILEGSRCSRVNGR GWRCCQPTLVGYSLCEHHLGKGRLRSMSASAK SRSSKKKSNSITKFALGRDEESDHTVDVREKKLE RMNSFGNKRVKLGVVKARSLSSLLGQTNAMKNN NIVADNHATMQQQEKISNLDYQ
661	sp_047290_czfo	-0.866726	-0.939562	-0.859587	G-box binding factor 1 (GBF1); CONTAINS InterPro DOMAIN/s: Basic-leucine zipper (bZIP) transcription factor (InterPro:IPR004827), G-box binding, MFMR (InterPro:IPR012900), bZIP transcription factor, bZIP-1 (InterPro:IPR011616); BEST Arabidopsis thaliana protein match is: basic region/leucine zipper transcription factor 16 (TAIR:AT2G35530.1); Has 3905 Blast hits to 3842 proteins in 356 species: Archae - 8; Bacteria - 123; Metazoa - 1421; Fungi - 442; Plants - 1419; Viruses - 137; Other Eukaryotes - 355 (source: NCBI BLink).	MWAGQHPMMPPYGTPVPYPALYPPGAVYAHPN MAMTPNPASVVETKGKNSDVKDQNSVKKSKGN SGLGNEKGVETAKATSGSGNEGASQSAESGSE GSSDASDDNNQQEHSATKKGSFDLMLEDGANA QHHSVNASAQGSVPGKPLGVAPNLNIGMDLWN SSAGGAKLRPSTGVASGVAPAMVGREGLMPDH QWIQDERELKKQKRKQSNRESARRSRLRKQAE CEVLQKRVESLNNENSNLREELKRLSAECEKLTS ENKSIKEELTQTCGPDSLQSLEENNASGEGDS
662	sp_047760_jdxq	-0.871505	-0.893404	-0.950371	SOUL-1; FUNCTIONS IN: binding; INVOLVED IN: N-terminal protein myristoylation, red or far-red light signaling pathway; LOCATED IN: plasma membrane, vacuole; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: SOUL haem-binding protein (InterPro:IPR006917); BEST Arabidopsis thaliana protein match is: SOUL heme-binding family protein (TAIR:AT3G10130.1); Has 1375 Blast hits to 1375 proteins in 156 species: Archae - 18; Bacteria - 186; Metazoa - 71; Fungi - 0; Plants - 173; Viruses - 0; Other Eukaryotes - 927 (source: NCBI BLink).	MGMILGKITVETPKYEVIKSTNEYEIRKYPASVIAE VTYDPSQMKGGNKDGGFTILANYIGAIGKPQNTT SEKVSMTAPVVTQSSNAAAAAEAIAMTAPVITKG REGEKEEVTMQFILPEKYKKAEEAPKPVDERVVI KQEGERKYGVVKFSGVASDGVVKDKVEGLRIAL DKDGFKVIGEYLLARYNPPFTLPFLRTNEVFLPVE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
663	sp_047840_szoo	-0.926036	-0.954788	-0.895022	general regulatory factor 8 (GRF8); FUNCTIONS IN: protein phosphorylated amino acid binding; INVOLVED IN: brassinosteroid mediated signaling pathway; LOCATED IN: in 6 components; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: 14-3-3 protein (InterPro:IPR000308); BEST Arabidopsis thaliana protein match is: G-box regulating factor 6 (TAIR:AT5G10450.1); Has 2678 Blast hits to 2669 proteins in 381 species: Archae - 0; Bacteria - 0; Metazoa - 1257; Fungi - 298; Plants - 764; Viruses - 0; Other Eukaryotes - 359 (source: NCBI BLink).	MAIPDNLTRDQYVYMAKLAEQAERYEEMAKFMD KLVVTTTSSGGELTVEERNLLSVAYKNVIGSLRA AWRIVSSIEQKEEGRKNDDHVVLVRDYRTKVESE LSDICACILKILDSNLIPTASVSESRVFYLKMKGDY HRYLAEFKSGEERKSAAEDTMEAYKNAQEIAVTD LPPTHPIRLGLALNFSVFYYEILNAADKACSMAKQ AFEEAIAELDTLGEESYKDSTLIMQLLRDNLTLWT SDQDQLDES
664	sp_048480_ixsg	-0.863868	#N/A	#N/A	DHFS-FPGS homolog C (DFC); CONTAINS InterPro DOMAIN/s: Mur ligase, central (InterPro:IPR013221), Folylpolyglutamate synthetase, conserved site (InterPro:IPR018109), Mur ligase, C-terminal (InterPro:IPR004101), Folylpolyglutamate synthetase (InterPro:IPR001645); BEST Arabidopsis thaliana protein match is: DHFS-FPGS homolog B (TAIR:AT5G05980.2); Has 7721 Blast hits to 7719 proteins in 2542 species: Archae - 43; Bacteria - 4847; Metazoa - 165; Fungi - 368; Plants - 133; Viruses - 0; Other Eukaryotes - 2165 (source: NCBI BLink).	MLHIQSSSFLRHGLFRISCFCERRWRLRGVAAR GVSCLPSCLNKHDPLAPGLRSGMIPTEIVENMTRI EVTEEYSWDLTHSRSYETVMEALSSLITCRKRGE KPSIGGKYTKLERMTMYLKILGLEENIAGLNIIHVA GTKGKGSTCAFSEAILRECGFKTGLFTSPHLIDVR ERFRLDGIDIAEEKFLQYFWDCWDQLRDNVTKD LPMPPLFQFLTLLAFKIFIDEKVDVGIFEVGLGGEK DSTNVIKEPVVCGISSLGMDHLDSLGDTLGKIALH KAGIFKPNVPAFTVPQLPEAMNVISQKAQLLRIPL EVVQPLTRGELNGLQLSLAGDHQFVNASLAVSL CKSWLRHTGNSEKVFTNDGCEADMPDAFLRGL SSARLPGRAQIVHDAIFRSSDTATGDLVYYLDGA HSPESMDACARWFSKMAKGDHKSTPLSSSKSS CENGLGEAVSKNGFTKQEKGNCDKISKQILLFNC MEVRNPQILLPKLIGTCASSGVHFSGALFVPSIST YNKVVSGYSITSNIPLRNLSWQFNLQRIWEKTVN GKDFDLDRNSMTECIKNESTDALPPLEFLHGDGY RDFPGDSFRSSAVVPSLSSAIKLLRDCARENPSIR LQVLVTGSLHLVGDVLRLLKR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
665	sp_049010_xxhz	#N/A	-0.851063	#N/A	G2484-1 protein (G2484-1); FUNCTIONS IN: RNA binding; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Agenet (InterPro:IPR008395), Tudor-like, plant (InterPro:IPR014002); BEST Arabidopsis thaliana protein match is: agenet domain-containing protein / bromo-adjacent homology (BAH) domain-containing protein (TAIR:AT1G68580.2); Has 6169 Blast hits to 4010 proteins in 631 species: Archae - 30; Bacteria - 1624; Metazoa - 1931; Fungi - 976; Plants - 400; Viruses - 26; Other Eukaryotes - 1182 (source: NCBI BLink).	MPMDYDDNDTTRHNIQLAGEGRSRSSSALRPYD FPKFDFDDTLQGHLRFDSLVETEVFLGIQSQEDN QWIEDFSRVSNGMEFNSTPADSCSISRRNNVWS EATSSESVEMLLKSVGQEEMMPGDSIITESNATH DQGIPNKNVESSIKLIERTDAYVVNSDAALPPDEL LEDCPASGRHIEKQLEDPSAFHAKVESASGISSN VNQNALCGNGGTLTYESNHSDHEERSGEKLEES QTSVGQTVNFGSGVDSLGFSSTTDNVTSNIGTLS TSKTSQQVMFNAFHESCSNDIPKSQVSIIAELDDT TILEGKANASSINIFTGSSSHGFVPESGRDCQEAS LDFVQESTNILVKGESELKSSEGVDKNTSLITSIQ TSRHEKVETALDDTAILEGKACVSSINIITGSSSR GSVPESGRDGREAILDFVRESANVLVKGESELKS SEGVDKNTSFTTSIQTSKHEKVVGMSNEVIEPFD DCPGRPSVDFASTEQDMDLTMQLPHKQDIEVND NMEDVSAIRLNHESMSEMPSLSIEDNGVPRGSA DCIRDTEVGNSSNISVQLSASDHQHNREVEGYH NECIVVGEEPGVAPDKDVPIQVIQENLVHAQADF SEIDICKPKEVDSEVSTDLGNIDSENVRTLKNDQR TQSAEVSEGDQGNNLIHKLQPQTSGEPESAINVS PFATTNASAALDDCPVIGQREPAAVGLFHSHNDD EEAGRKTSNDTNLGAEVDGASKVEDSVADDIVGI PVSEMAAKSEQPVDKIDDICSIGPSSVQAMNSLY SDHPQAHSDGNKKPSLKESADADLINVENHEKA MNKDEEKESAQIAEIDKTSCGSPTIISSSEPSQSE KDNQEEDRGYITNEVVKDIQSTSEDSKRKDLLGD DSSFTFKVNALPDTTEREPAKNWSPFPDVEKCK VVVETTPSGKGKTGRTPLKKTPRQNTQVSEGET GPGSSKGTSERKGRRKGTAKENAKNGNQSKGP TQENASMIMDKVSNMLSTPISNPIVQYGQMQAY GHTEANNKKPSGAVAVPTPSLPDLNASALNNST PSSTSYRQPFTDLQQVQLRAQIFVYGSLIQGTAP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
666	sp_049520_dnoz	#N/A	#N/A	-0.2663388	uncharacterized protein	MSFTNGPEKFFQYDMEELLENVPDELKKNPAFM KEVQSLADTKFPDDGPLNEHVTASPLVEVFAVSI HSYKSTLFNDGKPCEVYGEVKVTNVDTTDHLYK RVEEESEIVCLGGSLSLTGPTSSIVPYLTTKLNIHL RDRVRDIEIARGQIDLDFRTEDESEALKVGFVQG ECGFAAVLYTTLRSAVLANIVARFVNKYGFSNTD VAKQYFSTSNAAQKDCLSAVLFEKKPPSQYEYEY EYEYEYERVEHGTLIGLSRSLVVVPAYSSLVIEVD LWDHGTKLQFVNGTAEFRARRCGNGDCMNLQG QDCVLKLRVTWVHEDVCLSPPMIQNKRELAVTD KRKQKKTRTKFQKSPEPLRGHFPIIKGCASSVKP KRLCKFPMLEVFSVFIGQNNGKSAGLYGKIEVLC ALGRLRIFKREKDMPQQLSPFEKNSSTLNRCRGI PGTRLGIVVDFKDAHSPLVIQGSLWWCDFSLEYP KSWFDKRICSVIQGEYGYAAIHYTIFSAAVQAIVK ASISTNYSSDFPVQVYGSIVAQYTNVSYSTTYEN KYYRSVLFEKTQVSEITDFKLSKYVVAVPKDSSLII EVDLYLSSSIGESYGHGNDVENGTEHIRASAEFE DDGSGSGSGSGSGESFRTIKGQYYNIEFVVEWT DRLMTKE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
667	sp_049730_dsjn	#N/A	-0.879803	-0.920291	nuclear pore anchor (NUA); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: poly(A)+ mRNA export from nucleus, stamen development, negative regulation of protein sumoylation, negative regulation of flower development; LOCATED IN: cytosol, nuclear envelope, nucleolus, chloroplast; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide, MLP1/MLP2-like (InterPro:IPR012929); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MPLFLSDDEFERIKTAGDASAVAEKADSYIRDLYA NLETVKAEADAASITAEQTCSLLEHKYLSLSAEFS ALESRYSELDSNLQQRVSELAQVQSDKHQLHLK CIEKDGDIERLSIEVKELQKSKRQVVELVGVKDAE ISDKKATIKSYVDKITSLGSRSSDREARLSELEAEL VRCKANCDKLLQEKELIEKHNTWLNEELTIKVDSL LDQHRRSADIEADFSMKLADMEKQFNECTTSLS RHKEMVRELESKVESLHRELCSARDDASASEQQ FSAQISTISKLADLYKESSEEWSKKSGELEGVIKA LETHLSQVESDYKDKLEKETAARKEAEKVTVELK EKLVKCEAELESSRRANELDIIPLSSFTTKDLMYT SNTNDIVGDNLAIVPHIPAGVSGTALAASLLREG WSLAKMYTQYQEAADALRHEQLGRKQSQAILER VLLEIEEKASVILDEREEHDRMAEAYALMNQKLQ QSISEKVNLDKVILELKADLKKRERECTYAQKEIK DLQTQVTILLKECRDVQLRCGLDSNEYMNDDVA ALSDIDDESDTQKVISEHLVSVKDIKGLVEQNAKL RSLVRSFSNDIENKELQLKEQYEEELQKHKEEAA SKVAAVLERAEEQRQMIESLHSSVAMYRRLYEEE HRLRSSQPLSTEAPSTAGSKDFMLLLEGSQEAT KRVQEQAGERVRFLEEELASRKAEIISLRSERDKI ALEENFAREKLERFRKEFDHLREEMNGVLARNIE FSQLIVGYQKKLREGAESVSTAEELTRKLNMELS VLKREKDLLLNSEKRAQDEVQDLSKRVYRLQATL DTIQSTEQVNEAARAEEKRKQEEYIKKIEREWAE VKKELQEEKNKVQTLSLDHDHTIKDGMIQVENLR KELASALQVVKDAEVRAAVAEARYADLERKLKSS EVEVSEQDGRSIPLSSSAAEILGDVGFREEIERLK AESQASKDHMLQYKSIAEVNEAALKKMESAHDN FRIEAEQSKKSLEAEIQSLRDRVRELENEYTLKSE EATSVVVQKDEALASALGEINNLKEETSNKVSQL AALEVQFTMLKDDLEREHQRALAAQANYERQVIL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
668	sp_050810_qjgh	-0.895187	-0.946398	-0.905831	retinoblastoma-related 1 (RBR1); FUNCTIONS IN: transcription factor binding; INVOLVED IN: in 15 processes; LOCATED IN: cytosol; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Retinoblastoma-associated protein, B-box (InterPro:IPR002719), Cyclin-like (InterPro:IPR011028), Cyclin-related (InterPro:IPR013763), Retinoblastoma-associated protein, A-box (InterPro:IPR002720).	MEAKELQANFVHLSLLSKYYKRAYQEFFKSSEAD KDLVDASSSSSSSSGTISDYYKFGWLLFLALRVH AFSRFKDLVTCTNGLVSVLAILIIHIPARLRKFNVL DSPLFEKKTDRGVNLVSSLCQKYDTSEDELKKMI EKANELILEILKKKPCSESQCKTENLENIDPDDLTF FEDLLEDSSLPTCISMLEKDYETTICNKGDIDERV FVNVDDSLLASGSLSGGSVNISGTKRTFEAMASP TKTISSPLSPCRSPTHVNGARGGVNVRMMSTPV STAMTTAKWLRSYISPLPSKPSPQLEKYFMSCDR NISNEVVHRAHIIMEAIFPNSALGERNVAGNLPNS NLTDNIWAHQRRLEALKLYYRVLEAMCSAEAQLL HANNLTSLLTNERFHRCMLACSAELVLATHKTVT MLFPAVLEKTGITAFDLSKVIESFIRHEDSLPRELR RHLNSLEERLLESMVWEKGSSMYNSLIIAKPAMS AEINRFGLLAEPMPSLDAIAVHINMCSGGLPPLPS SIKHETATGQCPGSPKLSTFPSLPDMSPKKVAKN VYVSPLRSSKMDALISNSSKSYYACVGESTHEFQ SPSKDLTAINDRLNSTRKVRGALNFDTDAGLVSD SLVANSLYLQNGNCTTTTSSSVVPLKTEQPDS
669	sp_051030_zndn	#N/A	#N/A	-0.878069	uncharacterized protein	MAKSCGDEERVIILSKMLDEFDYLQSEICLDEKNS DRFESRMSEEDKLAAVGNGLPFTATFRDREYEF KYTRRKGYAIVEFDLARMVQHHPVTIGKRFLMIS MPNSCYQFFVIPENITGSVVVADSGKPIFRVGLIP NNVKM
670	sp_052180_skwk	-0.869396	#N/A	#N/A	pseudouridine synthase and archaeosine transglycosylase (PUA) domain-containing protein; FUNCTIONS IN: RNA binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Pseudouridine synthase/archaeosine transglycosylase-like (InterPro:IPR015947), Pseudouridine synthase/archaeosine transglycosylase (InterPro:IPR002478), Translation-associated RNA-binding, predicted (InterPro:IPR016437), Uncharacterised domain 2 (InterPro:IPR004521); BEST Arabidopsis thaliana protein match is: eukaryotic translation initiation factor SUI1 family protein (TAIR:AT1G71350.1); Has 885 Blast hits to 883 proteins in 291 species: Archae - 148; Bacteria - 0; Metazoa - 325; Fungi - 155; Plants - 86; Viruses - 0; Other Eukaryotes - 171 (source: NCBI BLink).	MFRKFTSEDVSAQNQVKASVQRRIRQSIAEEYPG LEPVLEDLLPKKSPLVVVKCQNHLNLVVVNNVPL FFNIRDGPYMPTLRLLHQYPNIMKKFQVDRGAIKF VFSGANIMCPGLTSPGGVLDEEVEIERPVAIMAE GKQHALAIGFTKMSAKNIKAINKGIGVDNLHYLND GLWKMETLD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
671	sp_052260_ihkm	-0.910166	-0.934383	-0.857464	RNA-binding KH domain-containing protein; FUNCTIONS IN: RNA binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: K Homology, type 1, subgroup (InterPro:IPR018111), K Homology (InterPro:IPR004087), K Homology, type 1 (InterPro:IPR004088); BEST Arabidopsis thaliana protein match is: RNA-binding KH domain-containing protein (TAIR:AT1G14170.3); Has 5625 Blast hits to 2559 proteins in 215 species: Archae - 0; Bacteria - 48; Metazoa - 3662; Fungi - 737; Plants - 967; Viruses - 0; Other Eukaryotes - 211 (source: NCBI BLink).	MAQQGQYPTIPTLQPYTQHHHQYIPPPHPPLQP PHPHHQPLNIPYPHQYHHHHPHQYADYASVPQL QHMPLPPIQYPTYPPPHYSDRLSSKRPREPVSDS AEVPSTKRQAVGNDILFRVVVPSGQIGKVIGKGG HRVQKIRDDTKSTIKIADPIARYEERVIIISSRENEA GGISDAEKALLQIASLILADDGTSAPAVSIAAATDP YAASGISGGHAVANSVKLLIAGSQAGGLIGVSGQ NIEKLRNSSGATIAVLAPNQLPLCASAHESDRMV QISGEVPAVMKAVEEIGVELRQNPPKQVISISPTY NLSIARAAQNYMDPNAAEYVTLDMVISETMVGGL IGRAGSNISRIRTESGAMIKVFGGKGENKHRQIQL IGSAQQVALAKQRVDEYIYTQLTQQPAAEQSV
672	sp_052480_wsyg	-0.910214	#N/A	#N/A	ubiquitin-specific protease 17 (UBP17); FUNCTIONS IN: cysteine-type endopeptidase activity, ubiquitin thiolesterase activity, zinc ion binding; INVOLVED IN: ubiquitin-dependent protein catabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, MYND-type (InterPro:IPR002893), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site (InterPro:IPR018200), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2 (InterPro:IPR001394); BEST Arabidopsis thaliana protein match is: ubiquitin-specific protease 16 (TAIR:AT4G24560.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MLEIPVFLSLALLIFTVVLRWTLKNEETKEVMVSR EAAMADFDAFQGDFSNSTVAMKAHECALCSSLT STRCSRCKAVRYCSSKCQISHWRQSHKYECRP QGPSASERDNFNGNAEMMEQHDTLSAEDFPRM DEKNTICQLNHDVLDSPVPRLSTVTIECSDIDTRW SSSGENGTPYSDINAPSMDQRKDKNNTLDTNTS MGLASNVVMNSAKKLNREKSRQNDSRKSRLSIS GLSSELRKDNSFHRAVSDNIHSSKSKELPFVDNE GLKHSQSIKHGKVKSMASKHSGNSLLLSSGSSS SQLPHNAFKSSLKKVVHQFNTPRQKKSSLHNSLL GISGNPNYKTVIQYDAFLSLYTSNDIELKPFGLTN CGNSCYANAVLQCLAYTRPLTMYLLQGIHSKEC WKDEWCFICEFENLMLKGREGYTPLSPIRISSNIS HLGHGREEDAHEFLRYAVDTMQSVCVELAGFVD PVREESTLVGMTFGGYLHSKIKCMKCHNISGKHE RMMDLTVEIDGDIRTLQQALRKFTASETLEGNNK YYCSSRCKSYEKAKKKLTIFQAPNILTIVLKRFQA GNFTKLKKRVQFPEILNITPYMSDSSDKYPQYSLY AVVVHLDVSSADSSGHYVCYVKNYHEEWFKIDD SVVISVDLERVLSEKAYMVLYARDTPRPPASGMS FHDEKPADYSGSYVMSKVRHSQSDVSVDPWVIQ HQSSKSGYVKDPSYVEEWITHSVESIPREDSSSE VSSLFSYSEEGTCSTASTKDSGSSGWFI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
673	sp_053060_crdw	#N/A	#N/A	0.6811039	pfkB-like carbohydrate kinase family protein; FUNCTIONS IN: kinase activity, ribokinase activity; INVOLVED IN: D-ribose metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Carbohydrate/purine kinase (InterPro:IPR011611), Ribokinase (InterPro:IPR002139), Carbohydrate/purine kinase, PfkB, conserved site (InterPro:IPR002173); Has 2896 Blast hits to 2893 proteins in 1061 species: Archae - 1; Bacteria - 2475; Metazoa - 103; Fungi - 59; Plants - 64; Viruses - 0; Other Eukaryotes - 194 (source: NCBI BLink).	MKKSNAMPIIIGGMVLDIHATASVPSTAKSTTPGQ NEFILGGVARNIAECLLKLGTKSFMISALGCDLPG DLLLEQWRSAGLLTEGIQRGLDITTASVCNMFDI NGDLTAGVASVESIEKYVTPQWIQQQFKSIVHSAP VMMVDANLSPEALKASCRMAADFDVPVWFEPV SVTKSERVAAIANYITFASPNEHELIAMANSLSRK NKFRPIERDQRKMNASIESLFKELKPAILVLLQKGI KYVLVTLGPDGAFLCSGSGPSGLRECLSSTKSFD KKREIFDKVSNRCPTHQYLGPTPYKKGSQLFVM HFPALPATVGRVSGAGDCLVGGMLASLCSGLDV MQSVAVGVAAAKTAIEVKTNVPDKYDLTTVADDA GAVYSAAKVVFQESML
674	sp_053320_mtgc	#N/A	#N/A	0.3533606	uncharacterized protein	SSLLWPPTAPTPSYGSTLTASLSILALTSLSTPRC FSYSTTITNNDLCECLHNRSADFWRWGREKP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
675	sp_053470_agdp	-0.320495	-0.83933	-0.824637	EMBRYO DEFECTIVE 2754 (EMB2754); FUNCTIONS IN: small GTPase regulator activity, binding; INVOLVED IN: embryo development ending in seed dormancy; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Citron-like (InterPro:IPR001180), Vacuolar sorting protein 39/Transforming growth factor beta receptor-associated domain 1 (InterPro:IPR019452), Armadillo-type fold (InterPro:IPR016024), Vacuolar sorting protein 39/Transforming growth factor beta receptor-associated domain 2 (InterPro:IPR019453); Has 871 Blast hits to 660 proteins in 186 species: Archae - 0; Bacteria - 0; Metazoa - 376; Fungi - 292; Plants - 87; Viruses - 0; Other Eukaryotes - 116 (source: NCBI BLink).	MVHSAYDSFELFNDCPTKIDAVESYGSKLLLGCS DGSLRIYAPESSSDASPPSDYRSTQLELLKEPYF QERIINGFAKRPITAMEVLSSKELLISLSESISVHKL SSFETVAVISKAKGANAFSWDDSRAFLCFSKQR RVYIFRHDGGRGFVEVKDYGVPDTVKSMSWCG ENICIGFRKEYMILNSSTGASSEVFSSGRISPPLV VPLPSGGFLLGKDNIGVFVDQHGKLGQEGRICW SEAPSQIVIHEPYAIGLLLPRHVEIRSLQQPFPLIQT VGLRNVRHVIRSNGAILVSLEKSINGLFPVPLGAQ IVQLTASGNFEEALSLCKLLPPEDSSLRAAKESSI HIRYAHYLFENGSYEEAMEHFGASQVDITYVLSL YPSITAPKSTALTEPQKVIDFGYDASDLSRCSSGI SDDMEASPRAQSDLEENATVESKIVSHNTLMALI KFLQKKRYHVIEKAAAEGTDEAVSDAFGTNFKSY DSNRSKKSNKGRGSIPISSSGGREAAILDTALLQA LILTGQSSAAMELLRGVNYCDLKISEEFLHKRNQY TVLLELYKCNSMHREALKLLLQLIEESKSEHSRPE LTQKFTPEMIIDYLKPLCGTDPMLVLEFAMLVLES CPSQTIELFLSGNIQADLVNSYLKQHAPNMQATY LEHMLSMNEDGISGNLQNEMVQIYLSEVLDWYK ELSSQQQWDETAYSPTRKKLLSALENISGYNPE GLLKRLPADALYEERAMLLGKLNQHELALSLYVH KLHVPELALSYCDRVYESQLNQSSRKSQSNIYLT LLQIYLNPRRSTRNFEKQNYNLVSTPNAEIKRVNS SASIRSKVGRSTRKIAEIEGAEDVRFSPSNTDSGK SDGDADEANYEGGSNIMLDEALDLLGRRWERIN GAQALKLLPRETKLKNLLSFLGPLLRKSSEAYRN FAVIKRLRESENLQVKDELYNKRKAVVKVTGDSM CSLCNKKIGTSVFAVYPNGTTLVHFVCFRDSQSM KAGRGSQLRKISQ
676	sp_054420_ycik	#N/A	#N/A	-0.822133	MAIGO 1 (MAG1); FUNCTIONS IN: hydrolase activity, acting on ester bonds, protein serine/threonine phosphatase activity; INVOLVED IN: endosome to lysosome transport, protein targeting to vacuole; LOCATED IN: microsome, retromer complex, membrane, multivesicular body; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Phosphodiesterase MJ0936 (InterPro:IPR000979); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MVLVLAIGDLHIPHRAPDLPAKFKSMLVPGKIQHI CTGNLCIKEVHDYLKTLCPDLHITRGEYDEDARY PETKTLTIGQFKLGLCHGHQVVPWGDLDSLAML QRQLDVDILVTGHTHQFTAYKHEGGVVINPGSAT GAYSSITYDVNPSFVLMDIDGLRVVYVYYELIDGE VKVDKIDFKKTTAPQSPH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
677	sp_054540_fhrj	-0.901186	-0.889249	-0.928636	methyl-CPG-binding domain 9 (MBD9); FUNCTIONS IN: methyl-CpG binding, DNA binding; INVOLVED IN: photoperiodism, flowering, secondary shoot formation, regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, PHD-type, conserved site (InterPro:IPR019786), Zinc finger, RING-type (InterPro:IPR001841), FY-rich, C-terminal (InterPro:IPR003889), Zinc finger, PHD-type (InterPro:IPR001965), FY-rich, N-terminal (InterPro:IPR003888), DNA-binding, integrase-type (InterPro:IPR016177), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Methyl-CpG DNA binding (InterPro:IPR001739), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: RING/FYVE/PHD-type zinc finger family protein (TAIR:AT1G77250.1); Has 6416 Blast hits to 3988 proteins in 224 species: Archae - 0; Bacteria - 0; Metazoa - 4085; Fungi - 602; Plants - 1260; Viruses - 0; Other Eukaryotes - 469 (source: NCBI BLink).	MKLNDSESPSIVAAQSSSTPSSPSSSHHRSFLAI DLNEVPSPSETLPSPSSSSSTFILSPTSLVRAIND NPPIPEGPPAEIPDDVTPCSGCGGPSPELVSDRL VCDGCERGFHFNCTGVLPVVREEWLCRYCVAE GVGSKRWRLGRKEDRGVRLIDMNASPPSDGDV ESLAAADAAAAAAGSSREVFGFM
678	sp_054640_zodi	-0.859161	#N/A	-0.877094	PLAC8 family protein; CONTAINS InterPro DOMAIN/s: Protein of unknown function Cys-rich (InterPro:IPR006461); BEST Arabidopsis thaliana protein match is: PLAC8 family protein (TAIR:AT3G18470.1); Has 865 Blast hits to 864 proteins in 99 species: Archae - 0; Bacteria - 0; Metazoa - 128; Fungi - 94; Plants - 626; Viruses - 0; Other Eukaryotes - 17 (source: NCBI BLink).	MEMIGIGSYVPPPYIPLSQSEEESEALITGNNSVQ ASASISNGTVQWSSGICACCDDSTSCCIGIFCPC YLFAKNAEFLGSETLVGPCMAHCMLWALASSLS CLYCVLLGLPGCLLSCYACNYRRILRTKYNLPEAP CGDFATHLCCHLCALCQEYREICERSGDPNYSD PKLPVITAPPVQKMEASPGE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
679	sp_055210_azxq	-0.357227	#N/A	-0.376538	NDR1/HIN1-like 8 (NHL8); FUNCTIONS IN: zinc ion binding; INVOLVED IN: biological_process unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ubiquitin-conjugating enzyme/RWD-like (InterPro:IPR016135), Zinc finger, RING-type, conserved site (InterPro:IPR017907), Zinc finger, C6HC-type (InterPro:IPR002867), Zinc finger, RING-type (InterPro:IPR001841), RWD (InterPro:IPR006575); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein (TAIR:AT3G14250.1); Has 3653 Blast hits to 3199 proteins in 223 species: Archae - 0; Bacteria - 0; Metazoa - 1736; Fungi - 604; Plants - 670; Viruses - 7; Other Eukaryotes - 636 (source: NCBI BLink).	MREKVRQFEAEEVELSEEQIQINDQVQEDELLAM ESIYGDNVYILDKRQGLRSFQIHINIELPDSYSVST KLKEDESLQTFKVQCLSPIILTCLLPRSYPSHYPP SFTISVQWLHESEISQLCSMLDLLWNEQPGQEIL YRWVEWLPSSSLQHLGFDEELSIGPYNVESVES VSDRRSFSGSISPDIDVASMKSYNSERQHENFLN GCHECCICFSEFAGTKFLRLPCDHFFCVKCMTTY AEVLVKEGTVNKLQCPNTKCGGMIPPGLLKRLLG GEEFDRWESLMLQKTLDAMPDVVYCPRCETACL EEKDHHAQCSKCFFSFCSLCRERRHVGLACMTP EMKLKSLEDQPIFSLLKESQRRKQRIVMNELLSV KEILRDAKQCPSFFFISRLTGMCDCHHSPAEPVIE PEQNMALQVRQILRELQVQMVHGNSRPCPSCK QFKAKVSLYIQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
680	sp_055380_ihrs	#N/A	-0.847992	#N/A	ROOT GROWTH DEFECTIVE 3 (RGD3); FUNCTIONS IN: binding, helicase activity, DNA binding, nucleic acid binding, ATP binding; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-like helical (InterPro:IPR011989), DEAD-like helicase, N-terminal (InterPro:IPR014001), Armadillo-type fold (InterPro:IPR016024), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021), SNF2-related (InterPro:IPR000330); BEST Arabidopsis thaliana protein match is: Homeotic gene regulator (TAIR:AT3G06010.1); Has 18805 Blast hits to 15736 proteins in 1976 species: Archae - 119; Bacteria - 5461; Metazoa - 3891; Fungi - 4072; Plants - 1633; Viruses - 163; Other Eukaryotes - 3466 (source: NCBI BLink).	MAQQAPPPQSQPQQQPQPQQSSSRLNRLLTLL DTGSTQQTRFAAARQIGDIARSYPQELNPLLRKV SQYLRSRNWDTRVAAAKAIGSIAENSKRVSIEQL VGIAARKVSEAGLGCPELKPKNEASISSFRSFDIN KVLEFGILLASGGQEYDVNDDNNAKNSKERLAR QKKNLRRRLGLDMCEQFMDVSDMIRDEDLVMFP MESSKSSSDYSRKRGGYRAFRSQRPSARELNLL KRKAKVNSNDVRCGVTVGAQKVVVASAGGSEP FVGGGGTGKGLVESATDEDSVEQEEDGYWQFH GYVEQLLVDMFDPVWEVRHGSIMALREILTYQG AYAGVSMNDPSSRGGQGTESANEKANLPVKRE REIDLNLEVSEDGSMESFKRVKHEGIEPSSVQMS VKAEQNGENASSPFVSEVNIGSVIMKSEPDALEA RSVSDGADSLMLGSVDTAEEKGIAGKVDMMDNL SGNLELINLVGSARNSWLKNCEFLQDCAICFLCV LSLDRFGDYVSDQVVAPVRETCAQALGAVLKYM APELVRETMNILLQMQERPEWEIRHGSLLGIKYLI AVRKEMMSDLLGFVLPACKCGLEDPDDDVRAVA ADALIPAASSIVANMNQLLHPIVMLLWDILLDLDDL SPSTSSVMNLLAEIYTHAGLTGNMLDTLTLQGPQ ELDLNEDLSYIDELELPASNENPYMLSALAPRLW PFMRHSISSVRYSAIRTLERLLEARYKRHITEPSA GSFWPSSILGDTLQIVFQNLLLESNEDILLCSERV WKLLLHCPVMDLVTAAKSHITSWLELATTPYGSE LNISKMFWPVAPPQKRKFVTAAKMRAAKMDNYA GRSGNVGAAKGTILEDCVDASGSSLKIVVGADSE KSVTRTRVVAAAALGVLASKWPESSINHITDPLW KDISSLSGVQRQISSMVLISWFKEVKNGEISPAQ GVNSALFEQFKNWLLDLLACTDPSLPTKDSSLPY AELSRTYSKMHSEASHLLHVVESCGMLKDLLLTT NIDLGNLGMDDAVDFALKVVLPSYDSSVDSPAEK LILDDVESSKQRLLTTTGYLKCVQSNLHVTVSSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
681	sp_055650_fznp	-0.851031	-0.877124	#N/A	RNI-like superfamily protein; FUNCTIONS IN: ubiquitin-protein ligase activity; INVOLVED IN: ubiquitin-dependent protein catabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Leucine-rich repeat, cysteine-containing subtype (InterPro:IPR006553); BEST Arabidopsis thaliana protein match is: F-box family protein (TAIR:AT5G27920.1); Has 15959 Blast hits to 6468 proteins in 357 species: Archae - 0; Bacteria - 920; Metazoa - 6194; Fungi - 1434; Plants - 4975; Viruses - 16; Other Eukaryotes - 2420 (source: NCBI BLink).	MAKTSSSSPSSPSSPSTTNVFDLISEELVFIILDYL NDNNPQDKKSPFSLVCKSFYSIESKHRRTLKPLRS EHLPSVLNRYPQISTLDLTLCPRITDFTLSVVATTA CVNTLRFVDLSRSKFFTGNGLAALVASCRSLTAL DLSNATELRDLAAAAIGEAKNLEKLWLGRCKLITD IGIGCIAVGCPKLKIIGLKWCLGIGDLGVELIAVKC KDLRTFDLSYLPDVISFVACAFSGVITSKCLPAILK LQYLEELILEGCFGIDDDSLTLLKQGSKSLKTLDIS SCQNVSDTGLSSLTSGIGCLQQLTLAYGSFVTLD LARSLQRLSKLQSIKLDGCLVTCSGLQAIGNYCA SLRELSLSKCSGVTDEGLSSLVKKQKELRKLDIT CCRKLTCVSIAHVAYSCTSLTSLRMESCTLVPRE AFVLIGRKCHLLEELDLTDNEIDDEGLKSISKCSNL SVLKLGICLNITDVGLSYIGDRCSKLVEIDLYRSMD ITDSGITAIARGCPSLEMINISYCQRITDASLISISK CAKLSTLECRGCPLMTSLGIAAIAVGCKHLTKLDI KKCYNINDFGVLPLAHFSQNLKEINMSCTSITDVG LLSLSSISCLQSMTILHLKGLSPRGLASALLACGG LTKVKLHSSFKTLLPQPLFEHLEARGCVFQWREK MFQDELEDSKSWKLQVAEMNNDS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
682	sp_056150_tfii	-0.895022	-0.694613	-0.861018	DNA-binding bromodomain-containing protein; CONTAINS InterPro DOMAIN/s: Bromodomain, conserved site (InterPro:IPR018359), Bromodomain (InterPro:IPR001487); BEST Arabidopsis thaliana protein match is: DNA-binding bromodomain-containing protein (TAIR:AT1G76380.2); Has 11096 Blast hits to 8233 proteins in 323 species: Archae - 0; Bacteria - 121; Metazoa - 4363; Fungi - 1457; Plants - 677; Viruses - 7; Other Eukaryotes - 4471 (source: NCBI BLink).	MGKVVVEKKKKKGRPSLLDLQKRSLKEQQEOQ QHQQSQQKKQQFQNNVKKIPNSKIPNPSPNSSA GIRRSTRNPNPNSDIDEELPSGVGKKKQQLKAV KLKLPNGGGEDDEEVNSDEGRENNSRSESGPE GDASPGLNLRKRKIDSISHGSGVEETDKSEKQGS AVATTKQASTQQECTLDDGSMTPLPDKSLLLFIL DRLQKKDTYGVFADPVDPEELPDYHEVISDPMDF STVRKKLANGAYATLEQFENDIFLICSNAMQYNA ADTIYFRQARSIQELAIRNFDNLRADSDDNEPEPE PKVVRRGRPPTKNLKTPPGRPPLERANSDFPEA TLATAGGTTNRSNNEVRKASHSEKFGSTDSFSR SLYGSRVGETYIGWSAERYERADDMSGSASKGI LMKHGKKHIVLDETRRNTYNSYSSACQGEPSVLT AFDRERKQLIALDCGGLSCYLARLSMISGYQPKH HNPEDELDTAYWRGPAFNAFCFLTESNRVSMDE EIFWGIREYLIPKLTNVGLHTEYGYARSLARFAAK VGPVAWKVASKKIERCLPTAVKFGPGWVGENEA PAERLMOQQSPASVGQKTPPPSSLPAPTDTSSA HTQGIADSSGDGSVLKTEGEQLSESQKPCNTSN VNEESGKPLCPPVAGNDATPLSSSNVPSTGAGP NLIAPEASASPPSSTVVNILNNGSGSMRPENPCPI NQTWSVHSRVNGHNNGSYGLNFAAQMGKMIGA GRPTMFNLQSNPSIVTVSRTETQPATDSNGLGE AKLENSCTIKPIDGLQKTGTNQKVGSTPLDLNVG FQSPGSPVCGGKADSVQTDLALQL
683	sp_056190_tguh	#N/A	-0.868938	#N/A	uncharacterized protein	MLANEPHLPSLPS
684	sp_056270_qrni	-0.881356	-0.922287	-0.927299	SULPHUR DEFICIENCY-INDUCED 1 (ATSDI1); FUNCTIONS IN: binding; INVOLVED IN: cellular response to sulfur starvation, regulation of sulfur utilization; LOCATED IN: cellular_component unknown; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide-like helical (InterPro:IPR011990), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT1G04770.1); Has 255 Blast hits to 246 proteins in 34 species: Archae - 2; Bacteria - 37; Metazoa - 0; Fungi - 0; Plants - 177; Viruses - 0; Other Eukaryotes - 39 (source: NCBI BLink).	MEKRINRSPGEKESLHAVVKVPSGDGPYVRAKH AQAINAGDRVDSALKDMAVVMKQLDRTEEAIEAI MSFRGLCSRNAQESLDNVLIDLYKKCGKVDEQID LLRRKLRLIYHGEAFNGKPTKTARSHGKKFQVSIK QETSRILGNLGWAYMQKTNYMAAEVVYRKAQMI DPDANKACNLGFCLIKQARYEEARSVLEEVVEAK LPGSEDIRSRNRAEELLDLESRQPPSLLSNIVCL DLEDEFVHGLEDLESKWGPLRPKRLPIFEEISNF KNQLAC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
685	sp_057280_dzei	-0.89778	-0.963297	-0.960354	pleiotropic regulatory locus 1 (PRL1); CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat, conserved site (InterPro:IPR019775), WD40 repeat (InterPro:IPR001680), G-protein beta WD-40 repeat, region (InterPro:IPR020472), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT3G16650.1); Has 84704 Blast hits to 34179 proteins in 878 species: Archae - 64; Bacteria - 9610; Metazoa - 34201; Fungi - 18356; Plants - 10881; Viruses - 0; Other Eukaryotes - 11592 (source: NCBI BLink).	MPGITASMEPVERQSLKKLSFKSLKRALDVFSPL HGQFPPPDSESKKIRLSHKICLEYGGIKPALAQPQ PTKQNQGRSSSAPSDPLALPAPEDAQRGGSHN ALVVAPAPQSRLNEVGPSKSNTLISIPGSSERFST SALMERIPSRWPRPVWRAPWKNYRVISGHLGW VRSLAFDPSNNWFCTGSADRTIKIWDVATGHLKL TLTGHIEQVRGLAVSQRHTYMFSAGDDKMVKCW DLEQNKVIRAYHGHLSGVYCLALHPTLDVLLTGG RDSVCRVWDIRSKTQIFALSGHDNTVCSVFSRPT DPQVVTGSHDSTIKFWDLRHAGKTMTTLTHHKK SVRAMALHPKENMFASASADNIKKFALPKGEFM HNMLSQQKTIINAMAVNEEGVLATAGDNGSLWF WDWTSGHNFQQSQTIVQPGSLDSEAGIYALSYD LTGSRLVTCEADKTIKMWKEDETATPETHPLNFK PPKDIRRF
686	sp_057580_pspt	-0.98208	-0.992154	-0.962676	GLIOMAS 41 (GAS41); INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: YEATS (InterPro:IPR005033); BEST Arabidopsis thaliana protein match is: TBP-associated factor 14 (TAIR:AT2G18000.2); Has 807 Blast hits to 807 proteins in 219 species: Archae - 0; Bacteria - 0; Metazoa - 350; Fungi - 284; Plants - 64; Viruses - 0; Other Eukaryotes - 109 (source: NCBI BLink).	MTQTPPSAPKDDSNRPEATASPINKPENSPEER ENKNLSRRVKDVEISVPIVYGTIAFWLGRKATES QSHRWTVYVRGGTNEDLSVVVKRVVFQLHPSFN NPIRVVDSPPFELSECGWGEFEIAISLYFHSDAG DRKLDVFHHLKLYPLEQNGPQSTKKPVVMETYD EVVFPDPSEQFLARVQNHPAVVVPRLPAGLELPS SASTEERRGSMKDHPLGHWFMNFSEADELLKLA AARQQVQAHIVKLRRQMSMVEGSSGTADPASAF
687	sp_057620_rwai	#N/A	-0.877831	#N/A	peptidylprolyl cis/trans isomerase, NIMA-interacting 1 (PIN1AT); FUNCTIONS IN: peptidyl-prolyl cis-trans isomerase activity; INVOLVED IN: regulation of cell cycle; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidyl-prolyl cis-trans isomerase, PpiC-type (InterPro:IPR000297); BEST Arabidopsis thaliana protein match is: FKBP-like peptidyl-prolyl cis-trans isomerase family protein (TAIR:AT1G26550.1); Has 6394 Blast hits to 6305 proteins in 1691 species: Archae - 14; Bacteria - 4880; Metazoa - 261; Fungi - 147; Plants - 126; Viruses - 0; Other Eukaryotes - 966 (source: NCBI BLink).	MKYGVYIYTHRSHTYLTKAIFLPRPEEENQTEKIE SSGRKRRKEASNDGDQGGGGGGVGERNREKE KRHHHHRHQEHKKSKKMSTTQQVRASHILIKHQ GSRRKASWKDPEGVVILNTTKESAISQLKSIRDDI LSGNANFEDVASRISDCSSAKRGGDLGPFGPRQ MQKPFEDATFALKVGEISDIVETDSGVHIIKRTG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
688	sp_057830_znmu	-0.89621	#N/A	-0.856198	fringe-related protein; FUNCTIONS IN: transferase activity, transferring glycosyl groups; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Malectin/receptor-like protein kinase (InterPro:IPR021720); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT5G24010.1).	MVNSSIIPAFSFLSLLPLLLLFLEVSSSFTPTDHHLI NCGSSAPSTIDADHRKFFGDESSLPHSVRLSASQ TIPLTDLEFKLSGDVSPLYNSVRIFRKPSKYTFRV KEKGAHLVRLHFHRFDSSNFDLSDAQFHVLADG YLLLNEFSGLTMKSLEIREYILWVDSDKLELSFKP LKKFNFGFVNAIEVISAPRDLILDSAQFVDSNGVIK IDGISRSAFETMFRVNVGGAKVTPFNDSLWRTW TTDDGYLESDHGSKEAHFGGKIKYQLGGASREV APDNVYNSARVVSASSSGVAKEYLTWGFPIVDG YKYLVRTHFCDIASPVLGSLYFNVYVNGDMSIENL DLSTITNSLASPYYADFVVNGENGVITVSIGHSKG SSSSGVDALLNGIEVFKMSNSMNSLDGMVSAKSI LRRCRGGNSSFLVLVVAAVGLLVAASVVLRKRLM AASDSVAWSPLPVDISEVSLKSTYL
689	sp_058230_ygyd	#N/A	-0.921798	-0.89689	Flavodoxin family protein; FUNCTIONS IN: oxidoreductase activity, FMN binding; INVOLVED IN: oxidation reduction; LOCATED IN: cellular_component unknown; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Oxidoreductase FAD/NAD(P)-binding (InterPro:IPR001433), Ferredoxin reductase-type FAD-binding domain (InterPro:IPR017927), Riboflavin synthase-like beta-barrel (InterPro:IPR017938), FAD-binding, type 1 (InterPro:IPR003097), Flavodoxin/nitric oxide synthase (InterPro:IPR008254), Flavoprotein pyridine nucleotide cytochrome reductase (InterPro:IPR001709); BEST Arabidopsis thaliana protein match is: P450 reductase 1 (TAIR:AT4G24520.1); Has 7053 Blast hits to 6598 proteins in 1587 species: Archae - 9; Bacteria - 3832; Metazoa - 1042; Fungi - 865; Plants - 492; Viruses - 0; Other Eukaryotes - 813 (source: NCBI BLink).	MEEKPKLLILYATQTGNAMDVAERIGREANRRGC PVSLLSIDQFDPGLLPNQRAVIFVVSTTGQGDTP DSLKVFWRHLLQRHLTRDWLGGVHHAVFGLGD SGYQKYNFVAKKLDKRLSDLGGTPIIERGLGDDQ HPSGYEGSLDPWMSLLWSSLYQINPSLFPIGVDV VPDMKLLDQSKVSVRYHVKDGASLIYSTTPDLES RASQIERARLMSPARFPSDKSRPHCWLKMTRNC LLTRAGSGRDVRHLEFGFHSIEVEYEVGDVVEML PSQNPDAVDAFIRRCNLDPESIITVQLRDSSNPDA DVDADAASGLPVRLRTFVESTMDVASASPRRYF FEVMSYFATAEHEKERLEYFASAEGRDDLYQYN QKERRTVLEVLEDFPSVQMPFEWLIQLIPPLKTR AFSISSSPIAHPNEVHLTMSVVSWVTPFKRKRYG LCSKWLAGLDPQQGVFVPVWFQKGLLPAPSPSL PLILIGPGTGCAPFRGFVAERAVQSQYGPTAPILF FFGCRNKEDDFLYSDFWLFHSQDGGVLSEEKG GGFYVAFSRDQPRKVYVQHKMQEHGQRVWDLL SNGASVYIAGSSTKMPSDVMSTFEEIISKESGVP KEAAARWLRALEKAGRYHVEAWS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
690	sp_058260_urcy	-0.8503866	#N/A	#N/A	RADIATION SENSITIVE23D (RAD23D); FUNCTIONS IN: damaged DNA binding, ubiquitin binding, proteasome binding; INVOLVED IN: response to cold, nucleotide-excision repair; LOCATED IN: nucleus; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Heat shock chaperonin-binding (InterPro:IPR006636), Ubiquitin-associated/translation elongation factor EF1B, N-terminal (InterPro:IPR000449), UV excision repair protein Rad23 (InterPro:IPR004806), Ubiquitin-associated/translation elongation factor EF1B, N-terminal, eukaryote (InterPro:IPR015940), Ubiquitin (InterPro:IPR000626), XPC-binding domain (InterPro:IPR015360), Ubiquitin supergroup (InterPro:IPR019955), UBA-like (InterPro:IPR009060); BEST Arabidopsis thaliana protein match is: Rad23 UV excision repair protein family (TAIR:AT3G02540.3); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKIFVKTLKGTNFEIEADSDDSVVDVKKSIEKSQG ADVYPASLQMLIFQGKVLKDATTLGENGVVDCSF LVVMISKNKPAAAGASSVSAAPAAPANQAQPQS STTPAAVAPASVAPQAQTQTATEPVVAAPAAPA ASGDENVYGQAASNLVAGSSLETTIQQILDMGG GSWDRDTVVRALRAAYNNPERAVEYLYNGIPEQ AEVPPVARTPAAGPTDLQVDDPLVQALNPLAEA GNPLAQAPQAAVPATGPNASPLDLFPQGLPNVG STAGGASSLDFLRNSQQFQALRAMVQANPQILQ PMLQELGKQNPQLMRLIQEHQADFLRLINEPVEG EGNLLSQFASAVPPTVSVTTEEREAIGRLEGMGF RRATVLQVYFACNKNEELAANYLLDHMDEFDNE EDEGLS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
691	sp_059170_rqsy	#N/A	0.8744324	#N/A	dentin sialophosphoprotein-related; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Occludin/RNA polymerase II elongation factor, ELL domain (InterPro:IPR010844); Has 17175 Blast hits to 8691 proteins in 788 species: Archae - 88; Bacteria - 6651; Metazoa - 4182; Fungi - 1520; Plants - 456; Viruses - 109; Other Eukaryotes - 4169 (source: NCBI BLink).	MYGGSHKFGHGGGARGGASANKHNSYFPPPPG RPSSAGRPSLSSRIPPSSSAATPPPATQSEESFS LVPGNPLSFTAIIRLAPDLIDEIKRVETQGGAARIK FDSNPNNPPGNVIDVGGKEYRFTWSHEMGGLC DIYEERQSCEDGNGLLVESGCSWRKLNVQRVLD ESTKKQVKKRSEEYEQKLKSRKAIVLDQCNPSM KSQMQALAAAEASPWKSKPKNKRRFEPSQVSG HPKPVKSTTVSSSTATAKGRQSTSPLPSPQNQS VPLTSAFVRRTATESQIVAENPTTNQTREKENAV SSGNVFPSQVSGGMLEAVVPRDQGAGISDLWN LLVTLLKESPKGMSIKALEKAVGDMTSNSLRKIEP ILKKIATCEPSGRYILKSGVELESLKRSYSDIGSSP EDNNQKQFVPEDNHGGVPAPDSGIAGKSPAQDV EEQVQILDEELNVIEKFDASQQFTDLFGDDEPLD NNEGQAMRTNDNKCESDSDSGSSSDSGSDSASR SRSPAGSGSASSSDSDGDDSSSHSKEASDVDV DIMSDDDKEPRNLLGTEHGLSTAPVPWSMPEVM PLQDGRNGDQDGYASEDVDIENDFPDDNVSQLA TAPRNLGIITEKLRTVSPDHDQNQPCNKADVITEH DVGDEFQRENHSFAKSGGKSKRGSNLIQFEHNG DHDNRFKENSAQPSMSRHILGSPPKHFADLINED IHESQNLQKSKQPNRDDSANAHLGVQNGYNYVV PSKSAVESLQSGRRPAARSYNTEPIDKAVKNPSN LGVELKFQKPNSSTQEGYSKHEDKDYGEAQDEE DGSNEKVNGKYKESPRGQKHSLRPESHHRKHS EWDKNSHSLGTSSRLRTGSSLKDSGKAEFDRSP ADRRLLHRELSDLELGELREPLPEEPAGNKRQFD RLGSFRQSETELGTSVSRNPELKSKSTSKPIVDS GNTSPPDARGAPNISYSSSRRRSLEPHVEDNYR PHGDLQSQLQQSRSGQAEVGSLPIKVVDHSRSG YTENVANQGTEAEGYGETHKKSQINSQREDSRQ EIGSRKGKGSKSRKSNASKDANDLTKKKKDASD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
692	sp_059430_zjkr	-0.939091	-0.938484	-0.958514	Transducin/WD40 repeat-like superfamily protein; FUNCTIONS IN: nucleotide binding; INVOLVED IN: biological_process unknown; LOCATED IN: cytosol; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40 repeat, conserved site (InterPro:IPR019775), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat (InterPro:IPR001680), WD40 repeat, subgroup (InterPro:IPR019781); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MLCGHAAPIVDLCISYPLLASLDEKVDENGGVGT SNQAALVSVCSDGVLCVWSRGSGHCRCRRELP PWVGSPSVIQALPTNKRYICVASCFTDISFASDP DYVDSTEGSEVTGDDESLSRKAPKCTIIIVDSYSL NIVQTVFHGHLSIGPLKFMALVESVDDSQKHSTFI IDSFGKSQCIPLIKDPNLEEGGAKLHRSSALENIAI LNNGSYDGELVLSVATYGQLLGLVYKTSCVFFLV ASATILGVISFEDSHHGGADCILSCVSGGMFFTSE SSDSTCMPESRKIHNITGLKFSVWDDRGSVIIYEL SYEDDTFKYEVINQIPATTCPINVKLSSKFVQLKQI FLRVESICFHNEEPSLWRPRVTIWSLHQQVDDR ENFGKICRLLGEGDFLVDWTDGSSYVINSEASKH NSGGIPATEEAPMTTMASALDQGVAGTLHKNDL NCTFIQKESIVSSTMVISLNAHIPYAIVYGYHSGQI ELVHFNFLAEEKDSQGPKKQFLGHTGPVLCLAA HQMVATSTERICSKILLSGSMDCTLRVWDIETCN AIMVMHHHVAPVRQIILPPLQSYHPWNDCFLSVG EDSCVALVSLETLRIERMFPGHPSCPLKVVWDG VRGYLACLSQNQSVASGTKDTLYIWDIKAGSRER VLRSTAAHSMFNNFCKGMDISNISGHAFSYNTSR SSLILPSTEQGCSLSPRLENYVPISSSEVNLMSSL NNSSTRSVKQNTGNSDKLPIRCSCPISGIAILSFD LSSLLSPDICGDVGGNFSNPPLNGSQTSINSERG TISNDSSEKGKFQDMMKDNQCIDTFRSCLLQFSL SFLHLWGVDFELDHLLTMEMNVKRYENVMLAPG LEGDGGSFTIAFPAFSSALQLWRSSSEFCAMRSL TILSLAQSLVSLSHSFSGASSALSAFYTRNIAEKV PSVKPPSLQLLVSFWQDEDEHVRLAARSLFHCA ASRAIPPPLCCLKTAENTKPVNSVDQVGQSELEN EKRAGTWASFLLPEMSTEVQGASLQVGQVDIRR WLESYELQDWISCVGGTSQDAMASHIIVAAALAI WYPSLIKPSLSMLVVHPLVKLVMAMNEKYSCTAA
693	sp_059670_uchn	-0.878004	#N/A	#N/A	uncharacterized protein	MDIQPESAQRITLGFMLVFKKFPDFDEQKVSLDL DADGLLGSVEGKYGILRLVP
694	sp_059830_wrtw	-0.977186	-0.976152	-0.999405	uncharacterized protein	MLDTPFGDDDNGASASGGQNGSFQMNPSMIKAI CNEVMKAMKGKQPANNSGAHSVAFANSAGISFE CSISCNLVFGDSWIIDSGATDHMTYNAKLFKSRK TLKIPIPICLPDAGSNGQKLGERNPILHIRHLGVFS KTLTAKLELPNYRAAAGQPPATVSPVSSATPGW LPMRVVQAHLL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
695	sp_059860_efns	-0.888544	-0.881104	-0.886321	Transducin family protein / WD-40 repeat family protein; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40 repeat, conserved site (InterPro:IPR019775), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat (InterPro:IPR001680), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT2G43770.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MEGGQYLGEISALCPLRVPSRFSSLPYLLAGSGS QILLFDVNLGRQIRSFCVFEGVRVHGIASGLSEEN ALSSSVTFKVVISGERRVKLFSLRFQLGSLDKPG VGTDLTFLQLLPRFSHWVLDFGFLEDGLSSSEEG FRYLAVGCNDNSVHIWDALTSSFIFEAICPERCLL YSMKIWGRDFRALLIASGTIFNEIVVWKLVQKNRK QPPTDVGQDPDSVFEDMHPLDCLHYEVDIISKLA GHQGSIFYLSWSSDGTKLVSVSDDRSARMWITN AEMDMTDDLTKLHGHYSAGPVLFGHGARIWDCC ISDSFIVTAGEDCSCRVWGLDGKELRVIKEHIGR GIWRCLYDPNTSLLVTAGFDSAIKVHCLHSSFSA VSGESIKGVKEREAEVFRVCIPNSSGQSMLMDS KSEYVRCLKFARKDMIYVATNHGYLYHAKLSESG TVQWTELIRITEEVPLICMDLFSRNLSEFEDWISV GDGRGIVKVIKVVFRDGAPQVAVRISWSAEMER QLLGVYWCKALGSGFIFTSNPRGLVKLWKLSDS VSHDAVRNCDVYLVASFMSPFGARIMCLDASVE EEALVCGDLRGNLILFPLSKRILHDSSILSIEMTPL TYFKGAHGISTVCSISIVRLTSSLVEICSTGGDGCI CYLEYNRDQQTVEFVGMRHAKELSLVQSFCSYR SSPDDSANISYAVGFSSTDFLIWNLLSETKVLQVK CGGWRRPHSYYVGQTPEICSGFAYVKDEIIYVHR YWIPDGEMKIFPRNLHLQFHGREIHSLCFIPDDVQ CATHFLHPKLSWVATGGEDGTVRLTRYNHGTGK WALSNLLGEHVGGSAVRSLCFVSKIYSAETKIIEA TDGTGGANNSLDKIENPCLLISVGAKRVLTSWLL KKSQRIEGDVSADGLEKKCDSHKSSPGHMSSMS FKWLSSDMPVKCVRTKKKAQSDGNLARPSDEV PGSEAVSSSMEIGLKQHGDAQDNDWRYLAVTSF LVNNAKSRFTVCFVVVACSDATLTLRALVLPYRY WFDVASLAPLSSPVLALQHAVAPLCQLYQGKMM TKNIYIIISGSTDGSIAFWDVTESIEAFMQRVSSLH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
696	sp_060250_reut	-0.857791	#N/A	#N/A	cysteine-rich RLK (RECEPTOR-like protein kinase) 41 (CRK41); FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 12 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Protein of unknown function DUF26 (InterPro:IPR002902), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: cysteine-rich RLK (RECEPTOR-like protein kinase) 29 (TAIR:AT4G21410.1); Has 118247 Blast hits to 116820 proteins in 4488 species: Archae - 94; Bacteria - 12773; Metazoa - 43795; Fungi - 10171; Plants - 33437; Viruses - 407; Other Eukaryotes - 17570 (source: NCBI BLink).	MIVRSFNTEGKVSLKLLVCLKVACQIISYVAFKRR NWXCKTRSSVEDLKWETRYKITLGIARGILYLRE NSCGAGPSDADKVVQTHATSS
697	sp_060590_znhr	-0.878835	#N/A	#N/A	BEST Arabidopsis thaliana protein match is: loricrin-related (TAIR:AT5G64550.1); Has 3596 Blast hits to 2004 proteins in 150 species: Archae - 0; Bacteria - 54; Metazoa - 1597; Fungi - 38; Plants - 513; Viruses - 9; Other Eukaryotes - 1385 (source: NCBI BLink).	MGTRNPGAAFCSDNLSNGYKNLGDSLQIAGCGS DHRADTILRLQPIGSNIVPHCPTSKGMKRKWGCR DGKTGEYVDPSLCLGLHRSSSSSDSKGSSATVC TAISSAKETEAESSMDLDLDFTLHLSSDKSPSPK KPTVSNFKRASMQLNLDLELSLSTSP-LESEITSIF GGLDTPQHGVVPLMVGATQFVDNESASSHWEV ENSPSPVPLSSGAGGISILKPIPRKVDLISATPDIS ASTFSVPKSSVTCTSGILQHPQRGSSSKLCQFTG CGKGARGASGLCIAHGGGRRCQRPGCNKGAEG RTALCKAHGGGRRCEFLGCTKSAEGRTDFCIAH GGGRRCGHEDCSRAARGKSGLCIRHGGGKRCQ RENCTRSAEGLSGLCISHGGGRRCHHLDCNKGA CGSTMFCKAHGGGKRCTVDGCTKGAEGSTSFC KGHGGGKRCSYSGCPKSVHGGTSFCVAHGGGK RCAVTGCTKSARGRTDFCVRHGGGKRCIFEGCG KSAQGSTDFCKAHGGGKRCAWGQLGSQYGNK VDDPCGSFARGKTGLCAMHGALVQDKRVHGCS TLGSLLQDPCSSDLGNMIGDVSAGIMNVDNMAM GPTGHYAAGWPPCDLMHAVEKESHSPPTPEGR VHGGSLLSLLASSSSVQAKQWP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
698	sp_060640_iwad	-0.887545	-0.903358	#N/A	XIPOTL 1 (XPL1); CONTAINS InterPro DOMAIN/s: Methyltransferase type 11 (InterPro:IPR013216); BEST Arabidopsis thaliana protein match is: S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (TAIR:AT1G48600.2); Has 22524 Blast hits to 22132 proteins in 2666 species: Archae - 571; Bacteria - 16701; Metazoa - 242; Fungi - 923; Plants - 574; Viruses - 10; Other Eukaryotes - 3503 (source: NCBI BLink).	MAASAMGVLQEREVFKKYWIEHSVDLTVEAMML DSQASDLDKVERPEVLSMLPPYEGKSVLELGAGI GRFTGELAEKASQVIALDFIESVIKKKNESINGHYK NVKFMCADVTSPSLNISPNSVDIIFSNWLLMYLSD EEVERLVERMLKWLKPGGYIFFRESCFHQSGDH KRKSNPTHYREPRFYTKIFKECHMQDDSGNSYE LSLIGCKCIGAYVKSKKNQNQISWLWQKVDSEDD KGFQRFLDSSQYKFNSILRYERVFGPGYVSTGGL ETTKEFVSKLDLKPGQKVLDVGCGIGGGDFYMA ENYDVEVVGIDLSINMISFALERSIGLKCAVEFEVA DCTKKDYPENSFDVIYSRDTILHIQDKPALFRSFH KWLKPGGKVLISDYCKSAGTPSAEFAAYIRQRGY DLHDVKAYGKMLKDAGFVEVIAENRTDQFIQVLQ KELDALEQEKDDFIDDFSEEDYNDIVDGWKAKLV RTTEGEQQWGLFIAKKM
699	sp_061230_uedu	-0.903431	-0.892794	-0.881335	Tetratricopeptide repeat (TPR)-like superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT1G80880.1); Has 37863 Blast hits to 13191 proteins in 288 species: Archae - 3; Bacteria - 35; Metazoa - 443; Fungi - 342; Plants - 35932; Viruses - 0; Other Eukaryotes - 1108 (source: NCBI BLink).	MWKFRRGFSLLNFSYDYCLGVRVNSMNLGRTH VPNFVTGYKLFSPVAENLKYPFSFSHKIQLFCSAT SDCLNLDSNSTELINVSDESGCGNDENDELSGE DDGIEAGKLGFIDEKLKHDLGVILEILRGSGPSQG ENSCDNPRLIGGAEVTNKLNQCGISLTSELVVEV LSRIRNDWELAIMFFLWAGKQPDYVHSVRQYHSI ISILAKRRKFDTAWSLIDEMRKAGIVNPKTLLIMIR RYAATHDVVNAINTFHGFKRFNFLVGVDDFQKLL SALTRYKNVQEAEQLLFSNRNTYQFNTKSFNIILN GWCNIVGSPREANRFWRLMIELGVKRDVFSFSSI ISCYSKLNRLKEVLKLFGQMKEMGILPDRKAYNSI MHALAKEGRAKEALNLLQIMEDKGIAPNVVTYNS LIKPLCKAKKRDEAQNVFDVMLSRGLKPSIRTFH AFLGIQRSGEEAFMLLDKMKELGCQPNKDSYLILI RKFCRWQQIDIVFTLWDQMKENGLNDDVTSHVA LIHGLFLNGKLEEAHKIYSEMKEKDMNPDVKTEE LIQTWLSNKQMELETSHVTSSETKLVDSDHVPG KTKSIRKQIGGGSDATEQPEFRKVVKERGFSFW E

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
700	sp_061800_gchx	#N/A	-0.8773819	-0.8708376	unknown protein; Has 34 Blast hits to 33 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 32; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MAKQTPTLFLEDWLRSSTSSSNVSNNSSKNSAN SSAQMIIQAWVDLRECHQKGAFHPRHLHSLKALV NSQASLYVADPQAKLLLSILSSPNISLPHDSPPLIF RLLYVWVRKSSKPSLALIDLAMEVIAHIFSDQFDA KKSKSLFAEGVLLVGAFCSVPSIAENTKTVCLKLV CRMLEEQYQLIGSVGKFIPDVLAGIGYALSSSGE VYFVRLLNFLLGNWYKIEVPHDISHALVILHLIEWV VFRLITSSSLLEKIKIFSCEVMENPKTNYAKYAVLM ATAGALRASNRPVKHGERLNIVTRFRIAAEAQIEV IASDLISKIATSSNIIIEDAEVSFILRCLLLAASRSGG LISSRGPLLLSLALGLLIEVLPLRRFYGMIVESQTS DNGVLIPVDVKNHLESILFREAGAATGVFCNLYAS ASEKDKSRVEDLIWSYCHDIYSGHRLVALLLQGR SDQLLNDMEKIAESVFLMVAVFAFAVTKKHKLTSK FSREMQMDCAVKILISFSCMEYFRHIQLPEYMEAI RSAATSVQENEIACVSFVESLPSYADLTNPQGLC HLQRGNYIWSKDEVQTSRIWFYLRVIPTCISHLPT NEFQKVVAATMFLYMGHPDNKVARASHLVFSSF VTFEEESSQKERMLLKEQLVFYYMRRSLEAYPGI TPFDGLGSGVSALVRHLPAGSPSIYYCIHSLVEKA NVLSREVTNLKGSEWWKNWQGDSEPCQKIIELL MQLISMVDIQVLPEMMKRLAQLVIQLPKDGQNVS LSYVCSQPTATNEACIEVSRSPNSLSLQNVTARL
701	sp_061930_gcjw	-0.8623132	-0.8623162	-0.8686588	HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription; LOCATED IN: intracellular, nucleus; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: High mobility group, superfamily (InterPro:IPR009071), High mobility group, HMG1/HMG2 (InterPro:IPR000910), ARID/BRIGHT DNA-binding domain (InterPro:IPR001606); BEST Arabidopsis thaliana protein match is: HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain (TAIR:AT1G76110.1); Has 3445 Blast hits to 3110 proteins in 301 species: Archae - 0; Bacteria - 5; Metazoa - 2269; Fungi - 308; Plants - 431; Viruses - 3; Other Eukaryotes - 429 (source: NCBI BLink).	MIDKGMSSGSNKHLDAAAAAAAAAAAAAAKYED VVSNRQLFMTTLENLHAAMSTKFMIPVIGGRDLD LHKLFVEVTSRGGIDKILAEKRWKDVTATFNFPST ATNASFVLRKYYYSLLQHYERIYYFKSQDCPPGV LQSLNATPVAVHGTAKMSPSPTPSHPANVQQNQ INSTEFVSQAPSGGYPIIGVIDGKFDSGYLVTVTIG KDKLKGVLYQTPPAKMNQVAQSSGVSTVKTDST RATSGVPRRRKRRKKSELRKRDPTHPKPNRSGY NFFFAEQHARLKPLHPGKDREISRMIGELWNKLK DNEKAVYQAKAIKDKERYKTEMEDYKERLRTNQ VISDAIPIQQRFPDIDVEMAEVDGKDETERDSQES PENESDSSKSETDEDEDDDDEEFCDADTPPAPD NAGTSEASVAAGAGTGTGTGEFEVVKGAEKGG FKESHVNVGDEAIFST

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
702	sp_062540_czgd	-0.876649	-0.854482	-0.936602	Surfeit locus protein 2 (SURF2); CONTAINS InterPro DOMAIN/s: Surfeit locus 2 (InterPro:IPR008833); BEST Arabidopsis thaliana protein match is: Surfeit locus protein 2 (SURF2) (TAIR:AT5G14440.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MATSEEKPTKNNIKEGSNLLGSPTFKQLDNGRFK CVETGHELPAHAQDSYSQSKRCRIGLIDFSLAHK KPPLNMFKQDPLNSSKLICKLTDDTINKSEEHIWK HINGKRFFNKLEQMEAGKLSSSGDVEEKGEKQG KAGLKKEENGAEKKKKKQKKMKETKKEEIVSEM RDSPGLKRKENGAEKKKEKQKKMKETKKEEIVS DSDLEGLEFWMPPAGERWDFDDGGDRWGSDS ESGDQHDEDGEKGETSI
703	sp_062800_jhai	-0.976101	-0.932305	-0.93652	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity, nucleic acid binding, ATP-dependent helicase activity, ATP binding; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), RNA helicase, ATP-dependent, DEAD-box, conserved site (InterPro:IPR000629), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT3G09620.1); Has 370658 Blast hits to 166919 proteins in 4473 species: Archae - 2452; Bacteria - 63379; Metazoa - 145404; Fungi - 45975; Plants - 23206; Viruses - 1711; Other Eukaryotes - 88531 (source: NCBI BLink).	MGESEDVKSKSRKEDSETKEESKRSRVREHDKE RDKDKGRNGEKERRRDREKERDRESRRSERDR SVDDKHRERERERDKGKHRDKHRDRDREKERK KDQKDRDGDREDKEKERDKEKAKEKEREREER ERERERERERERERERERERERERERERERERE RERERRERERERERERERERERERERAERDK EKEKERERERRRREAASDDSGDERKDRDRKRR RRDDENKERERERSKSRDREKSRSHKHRSGSV DSQKKKGDDEPDRSQEIMERKEQDMEEEQRRL DDEMEKRRRRVQEWQELKRKKEEDERKQLGDA NDEEPKAGKNWTLEGESDDEEAATEGQSDADM DVDRDFKPPTEAGDPMAIDVENENKLENVLPNG GADVPEEDDIDPLDAFMNSMVLPEVEKLSSSVLD TENAELNGKEEIVEQSNVEVSKKAPKKYLGRIMP EEDSDPDYEDQEDGGTNLEDEDDDEFVKRVKKT KAEKLALVDHSKIDYIDVRKNFYIEVKDIARMALED VDSYRTELELKVHGKDVPKPIKTWHQTGLTTKIL EVIRKMNFEKPMPIQAQALPVIMSGRDCIGIAKTG SGKTLAFVLPMLRHIKDQPPLMSGDGPIGLIMAPT RELVQQIHNDIKKFAKVMNLTSVPVYGGSGVAQ QIGDLKKGAEIVVCTPGRMIDILCTSSGKITNLRRV TFLVMDEADRMFDMGFEPQITRIVQNIRPDRQTV LFSATFPRQVEVLAKKVLNKPVEIQVGGRSVVNS DITQMVEMRPENERFLRLLELLGEWYEKGKILVF VQSQDKCDALFRDLLKSGYPCLSLHGGKDQGDR ESTLADFKSNVCNLLIATSIAARGLDVKDLELVLN YDAPNHYEDYVHRVGRTGRAGRKGHAITFISEDE AKYAPDLVKALELSEQPVPQDLKLLADGFAVKVK EGLEQAHGTGFGGSGFKFDEQEDEARKAAKKA QAKEYGFEFDKSDSEDEDDEVRKRSDVTVPTPV LGGQGQPDTTGHNVSGQGTIGSVPGSILSSINSS DNPALAVAVAAAMKLQQNLAKLQQDAMPEHYEA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
704	sp_062990_kkia	-0.883878	-0.848364	-0.914864	PCF11P-similar protein 4 (PCFS4); FUNCTIONS IN: zinc ion binding; INVOLVED IN: positive regulation of flower development, mRNA polyadenylation; LOCATED IN: nucleus, cytoplasm; EXPRESSED IN: shoot apical meristem, sepal, gynoecium, stamen, leaf; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF618 (InterPro:IPR006903), RNA polymerase II, large subunit, CTD (InterPro:IPR006569), Zinc finger, C2H2-type (InterPro:IPR007087), ENTH/VHS (InterPro:IPR008942); BEST Arabidopsis thaliana protein match is: Pre-mRNA cleavage complex II (TAIR:AT1G66500.1); Has 656 Blast hits to 406 proteins in 169 species: Archae - 0; Bacteria - 0; Metazoa - 208; Fungi - 279; Plants - 120; Viruses - 0; Other Eukaryotes - 49 (source: NCBI BLink).	MPSDTVMKPSVPSPILDRFRFLLKEREAEFRGSD DDEDDLVCIPPPSMEEIVRFYDDVLSELTFNSKPII TDLTIIAGEQREHAEGIADAICARVLEVPVDQKMP SLYLLDSIVKNIGRDYARHFASRLPEVFCEVYRQV PPNLYSSMRHLFRTWSTVFPSPVLGMIEDELQFS PRANQPSSNLASMRASESPRPSHGIHVNPKYLEI RRQFDNSSINNNIQPGRGASKSHGQKHSIEYEEY DYYQGEGTNSHLDAQRFNSTGYGGRTSFPSSTD QLLPSASGSVRPLSPSADEFFMSNSPGRLFEGV SPSHPGLNYGRRRGRDEENGDWQKRQRSDDS YHYSLNNGAEHDRPRALIDAYGQDDGQRSLKKN PLCMDRLIVNGGGSWQNTEEEEFDWEDMSPTL LADGKTGNVTTSTVPSRGFKAIPGVGSLGSRPFN PNGRKAVEDGGHLSSNRPVSRSEMNLSQGAWK LPVPVAGRENTSQLTLPVNGVPSSSSDGKPLSA DHYAPAPMQLGRISNSGFNSYEARAPSSAVAPV STSSWPSTSSHNFASLPPLYLQQKQPRASFDFR NQSYSILNQGNQQFDSDDKRNLIPPKPGQRPQL SFTPLQYLTQQPGSLQQQLHLPPEPRHNAVSLP PQTILPSGLGYTPLGHGVAASNIVRNPVPGSHPY APFQNVSNGYMQFQSPMAPLPLVPRPHSQFLPL PQSSGPLANQAPTAAISGLLNSLVAQGLFPSNQT PAQDSLGVEFNPDLLKVRHESAIRALTYEELPRQC TTCGLRFKTQDEHSKHMDWHVTKNRISKNRKHN PSRKWFVSVNMWLSGAEALGTDSAPGFLPTES VVEKKDDEEMAVPADEDQNSCALCGEPFEDFYS DETEEWMYRGAVYLNAPLGSTVGMDRSQLGPIV HAKCRSDSNDAPT
705	sp_063110_xnqj	-0.898876	-0.883423	-0.907733	uncharacterized protein	MGELVDGGEMFKRIFVMFAFSVFLAPIANRTADL GLVKVLEDVDEIKNLDWCGYVHVRLCRAIRKYKT SGCQGNVCGCLWVLQVVYFHRLQFRGIVESCSL PLMKHWSGTKIHERLILEKDSACFGSGILNTVTYP VCQKLGFEDGFAKICGRNDAANDDENHIRFILMV NCQIQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
706	sp_063170_tysi	#N/A	-0.9244372	-0.9838581	ankyrin repeat family protein; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: Ankyrin repeat family protein (TAIR:AT4G03500.1); Has 38664 Blast hits to 17395 proteins in 690 species: Archae - 39; Bacteria - 2692; Metazoa - 19825; Fungi - 3120; Plants - 3302; Viruses - 203; Other Eukaryotes - 9483 (source: NCBI BLink).	MEDDITMSLELYAAAVNGDMSIFPNQGSGEEATS GTSRSNQVDIYFLTVTLEERNTILHVAARTGHNLS FIAEALKRFPILIFQTNSKGETALHVSARQGNKEIT KLLVTFYRDVEGAAAGKNGSMPLWRVKNSEGD TPLHTAIKRGKIQVALFLISVDNSLAISVNNSRETP LHLAAKICSRIGALHVSARQGNKEITKLLVTFYRD AEAAAAGQNGSMPLWRVKNSEGDTPLHTAIKRG KIQVALFLISVDNSLAISVNNSRETPLHLAAKICSRI GASKRDIDELTPLLRASCCKSNLPFILVLINHNPQ FAEQLDLNGRNILHLIQIENYEDGKRALEIPEINAL KDEQDFNGETPLHLAIKNSDFMKVRVLMECSADF TIKSSYGISPWDLIQSQDNFANKMKDETDIAMNG MMKKELYKASITGNVQFLSRHDDIYLLSRSMEGN NILHTAMSPESTQAIEFIKTILQRIPILLCQTNANGD TPLHFAARARTSAAHIELTISCKNSFQKVLDMGAS FLYSTPPWRLRNSEGNTPLHEALRAGNSTCARF LIEVDDEVASFVNNYNETPLHVYARYLQIPGPLW GNQISERPLPQFCNRNAERISTQLSELDLLLEANI SAAFLRDMEGLTPLLRAAQCNYNIENIRKILDKCP SCAEIRDLSGCTLLHLIPLKDKHEVISHIHKLSNAR DHEGNTPLHTAINNRQFNKVIEGFT
707	sp_063330_mzmr	#N/A	-0.91023	-0.9263827	IRREGULAR XYLEM 12 (IRX12); FUNCTIONS IN: laccase activity; INVOLVED IN: secondary cell wall biogenesis; LOCATED IN: endomembrane system, apoplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Multicopper oxidase, type 3 (InterPro:IPR011707), Laccase (InterPro:IPR017761), Multicopper oxidase, type 2 (InterPro:IPR011706), Cupredoxin (InterPro:IPR008972), Multicopper oxidase, copper-binding site (InterPro:IPR002355), Multicopper oxidase, type 1 (InterPro:IPR001117); BEST Arabidopsis thaliana protein match is: laccase 10 (TAIR:AT5G01190.1); Has 9472 Blast hits to 8326 proteins in 1422 species: Archae - 45; Bacteria - 3703; Metazoa - 467; Fungi - 3359; Plants - 1600; Viruses - 0; Other Eukaryotes - 298 (source: NCBI BLink).	MSSFMDSPVAVDNSTNYGLLRYKPSLGGSPLVL TAMPSLNATQAQRNFIHSLRSLNSKKYPAKVPLSI DHSLFFTMGVGVNPCATCLNGSKLVGNINNVTFV MPTISLLQAHYYNISGVFTDDFPGKPLISFDYTTV NYTGSLATMNGTKVYRLPFNATVQVVLQDTSIIA PESHPIHLHGTNFFVVGTGIGNYDATKDSINFNLV DPVERNTIDVPTGGWTVIRFLADNPGVWFMHCH LEVHTSWGLKMAFLIENGDGPNESIMAPPKDLPK C

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
708	sp_064020_rckn	-0.932405	-0.87684	-0.890122	casein kinase alpha 1 (CKA1); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: casein kinase II, alpha chain 2 (TAIR:AT3G50000.1); Has 85531 Blast hits to 84684 proteins in 2530 species: Archae - 70; Bacteria - 9546; Metazoa - 32860; Fungi - 11294; Plants - 14699; Viruses - 299; Other Eukaryotes - 16763 (source: NCBI BLink).	MTWKYQITYTTNPFSSSLSLLFFRLPLLCALLALR APVALPPILRPPLSSSTLNYSFSAAINPDYFEISAA GKSSMSKSRVYTDVNVLRPREYWDYESLTVQW GDQDDYEVVRKIGRGKYSEVFEGINVNSSERCVI KILKPVKKKKIKREIKILQNLCGGPNIIKLLDIVRDQ HSKTPSLVFEFVNSTDFKVLYPTLSDYDIRYYIYEL LKALDFCHSQGIMHRDVKPHNVMIDHDLRKLRLI DWGLAEFYHPGKEYNVRVASRYFKGPELLVDLQ DYDYSLDMWSLGCMFAGMIFRKEPFFYGHDNHD QLVKIAKVLGTDELNAYLNKYNLELDPQLDALVG RHSRKPWSRFVNQDNQHLVSPEAIDFLDKLLRY DHQDRLTAKEAMAHPYFSQVRAAESSRTRTQ
709	sp_064530_dnim	-0.961435	-0.905865	-0.933187	eukaryotic translation initiation factor 4A1 (EIF4A1); FUNCTIONS IN: protein binding, ATP-dependent helicase activity, translation initiation factor activity; INVOLVED IN: response to cadmium ion, translational initiation; LOCATED IN: cytosol, nucleolus, cell wall, membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, ATP-dependent, DEAD-box, conserved site (InterPro:IPR000629), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: eif4a-2 (TAIR:AT1G54270.1); Has 49714 Blast hits to 49117 proteins in 3142 species: Archae - 779; Bacteria - 27442; Metazoa - 6117; Fungi - 4773; Plants - 2624; Viruses - 17; Other Eukaryotes - 7962 (source: NCBI BLink).	MRLGLFMAGAAPEGSQFDARQYDEKMTVLLNEE GGTEFFTSYDEVHDSFDKMGLKENLLRGIYAYGF EKPSAIQQRGIVPFCKGLDVIQQAQSGTGKTATF CSGILQQLDYELLECQALVLAPTRELAQQIEKVM RALGDYLGVKVHACVGGTSVREDQRILAAGVHV VVGTPGRVFDMLRRQSLRANNIKMFVLDEADEM LSRGFKDQIYDIFQQLPPKIQVGVFSATMPPEALE ITRKFMNKPVRILVKRDELTLEGIKQFYVNVDKEE WKLDTLCDLYETLAITQSVIFVNTRRKVDWLTDQ MRSRDHTVSATHGDMDQNTRDIIMREFRSGSSR VLITTDLLARGIDVQQVSLVINYDLPTQPENYLHRI GRSGRFGRKGVSINFVTGADDRMLFDIQKFYNV VIEELPANVADLI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
710	sp_064900_adyo	#N/A	0.837647	#N/A	RAB GTPase homolog G3F (RAB7B); FUNCTIONS IN: GTP binding; INVOLVED IN: protein transport, small GTPase mediated signal transduction; LOCATED IN: plasma membrane, nucleus; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ras GTPase (InterPro:IPR001806), Small GTP-binding protein (InterPro:IPR005225), Small GTPase (InterPro:IPR020851), Ras (InterPro:IPR013753), Ras small GTPase, Rab type (InterPro:IPR003579); BEST Arabidopsis thaliana protein match is: RAB GTPase homolog G3E (TAIR:AT1G49300.2); Has 26432 Blast hits to 26399 proteins in 759 species: Archae - 38; Bacteria - 171; Metazoa - 13566; Fungi - 4011; Plants - 2887; Viruses - 20; Other Eukaryotes - 5739 (source: NCBI BLink).	MPSRRRTLLKVIILGDSGVGKTSLMNQYVNKKFS NQYKATIGADFLTKEVQFDDRLFTLQIWDTAGQE RFQSLGVAFYRGADCCVLVYDVNSMKSFDNLNN WREEFLIQASPSDPDNFPFVVIGNKIDVDGGNSR VVPEKKARAWCASKGNIPYFETSAKEGLNVEEA FQCIAKNALKSGEEEEIYLPDTIDVGNGNQQRST GCEC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
711	sp_064910_gogj	#N/A	#N/A	0.884433	2-oxoglutarate dehydrogenase, E1 component; FUNCTIONS IN: oxoglutarate dehydrogenase (succinyl-transferring) activity, cobalt ion binding, zinc ion binding; INVOLVED IN: glycolysis, metabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: 2-oxoglutarate dehydrogenase, E1 component (InterPro:IPR011603), Dehydrogenase, E1 component (InterPro:IPR001017), Transketolase-like, pyrimidine-binding domain (InterPro:IPR005475); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate dehydrogenase, E1 component (TAIR:AT5G65750.1); Has 10999 Blast hits to 10962 proteins in 1954 species: Archae - 31; Bacteria - 4409; Metazoa - 546; Fungi - 299; Plants - 167; Viruses - 0; Other Eukaryotes - 5547 (source: NCBI BLink).	MTWFRAGSNVAKLAVRRTVSQGRSYVTRTRVID SQSRFVNSQSRFFHSTVLKPKAAPVPRPVPLSRL TDSFLDGTSSVYLEELQRAWEADPDSVDESWD NFFRNFVGQAATSPGISGQTIQESMRLLLLVRAY QVNGHFKAKLDPLSLEQRPVPGDLDLDNYGFTE ADLDREFFLGVWRMAGFLSENRPIQTLRNILNRL EQAYCGSIGFEYMHIADREKCNWLRDRIETPTPM EYNKERRHVMLDRLIWSTQFESFLATKWTTAKR FGLEGCETLIPGMKEMFDRSADLGVENIVIGMSH RGRLNVLGNVVRKPLRQIFSEFSGGIKPIDEVGLY TGTGDVKYHLGTSYDRPTRGGKRIHLSLVANPS HLEAVDPVVVGKTRAKQYHSNDVDRTKNMAILIH GDGSFAGQGVVYETLHLSALPNYTTGGTIHIVVN NQVAFTTDPTAGRSSQYCTDVAKALDAPIFHVNG DDLEAVVHVCELAAEWRQTFHSDVVVDIVCYRR FGHNEIDEPSFTQPKMYKIIRNHPSSLDIYKKKVL ESSEMSKEEVDQIQKKVMTILNEEFDASKEYVPQ RRDWLSAYWLGFKSPEQLSRIRNTGVKPEILKNV GKAITTLPETFKPHKAVKRIYEQRAQMIESGEGID WAVGEALAFATLLVEGNHVRLSGQDVERGTFSH RHSVLHDQESGEKYCPLGHVMINQNDEMFTVSN SSLSEFGVLGFEMGYSMENPNSLVIWEAQFGDF SNGAQVIFDQFLSSGESKWLRQTGLVVLLPHGY DGQGPEHSSARLERFLQMSDDNPYAIPEMEPTL RKQIQECNWQVVNVTTPANYFHVLRRQIHREFR KPLIVMAPKNLLRHKECKSNLSEFDDVKGHAGFD KQGTRFKRLIKDQNMHSDREEGIRRLVLCSGKVY YELDEQRKKENGSDVAICRVEQLCPFPYDLIQRE LKRYPNAEVVWCQEEPMNMGAYSYIAPRLSTAM KSVNRGNADDIKYVGRPPSAATATGFISVHGKEQ TGLVQKALQPEPIQFP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
712	sp_064930_ncsf	#N/A	-0.803408	-0.855722	CHB3; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity, zinc ion binding; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: SANT, DNA-binding (InterPro:IPR001005), Homeodomain-like (InterPro:IPR009057), Myb, DNA-binding (InterPro:IPR014778), SWIRM (InterPro:IPR007526), SANT, eukarya (InterPro:IPR017884), Zinc finger, ZZ-type (InterPro:IPR000433); BEST Arabidopsis thaliana protein match is: SWITCH/sucrose nonfermenting 3C (TAIR:AT1G21700.1); Has 17747 Blast hits to 11817 proteins in 958 species: Archae - 45; Bacteria - 1957; Metazoa - 7032; Fungi - 2514; Plants - 988; Viruses - 139; Other Eukaryotes - 5072 (source: NCBI BLink).	MEEKRQEPGSPSPAATTSGGSSINPSDSPATEPP APRRRNQKRKSTNTTSSSAPPKRHAREKANAAA AAASGGFFSLLPPIHNGPLTRARQLSDNNAAIFSA QNAIKNETSMAAVNDSAASEGGEEAKLAKEECE KLQAVIDAEFDEIRNRDSGVHVVPVAAGWFSWE NIHSIEKQTLCSFFNGKVENRTPDVYKEIRTWIMK RFHANPTTNIELKDLTELSIGDLDARQEIMEFLDH WGLINYHPFPQNEPNIDADPNTDADKAEQTDSLI EKLYGFESKQSSAQLVPRANVSAAAVPSGLFPE SITEELVKQEGPAVEYHCNSCSADCSRKRYHCQ KQADFDLCSECYNNGKFGSGMCPSDFILMEPAE ASGGTGGKWTDQETLLLLEAALELFKENWNEIAE HVATKTKAQCILHFLQMPIEDTFLDDDDKKDGAL ENGEPNSTSNEQPVPKGDSEPLEGDNGKDNVG ATETGSGKDHSEPSASKTSKDNEEPSETKTDND AAQLVSAQTETSNSEDDLNGKVVQGRGDDIIIKA LKEAFHVAGCPLTPEDELSFAEAGNSVMALAAFL TQLVEPGLATASSCISLKTISQSSPGSQLAARHCF ILEDPPDVDKEHTVSKRMATENIDPNTSEIAKEES KNDKEIIQEEKSTPLLDEHEKKESNASDDPRLPP GEKDTCADNSQIMEKPDEDMPQVTEKPDEDMP QVAEKPDEDMPQVVEKPDEKSQLVDKSGEKSQ PVDKPDEKSKLVENPDKKSQLVEKLVEKSQLME KLVEKSQLMEKPDEDLSQVTEKPDAAKPRVTEK PDEEKPQVTEKTDAEKPQVTEKTDEEKPQVTEK PDDEKPQVTEKPTEEKSQAPENPDEEKVMVEEK HDEVTNVEQAGMKSVNKPQNCSLQNEQSITAPK DQVTSMSEVMPPPVSENDSEIFKTTEENVQCTD VPKDSDIELSSVVSDKVEQQEIAGAVPMVRVASE ADKTVIDKITDPKSTKTARTKDEDNVAKLKRAALA TLSAAAAKAKLLADQEEEEIKQLAXPAPPPPRPLR KLEIKLSYFTEMDNAIMRVREQLERSRQKLFNER

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
713	sp_065040_fsoy	#N/A	-0.862737	-0.849503	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT3G52660.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MAEGSEIDERVDLDDDNYMEETDEDVVDGIEDE PEEGLGDDKAEEQYEELRTDGSGKEQSSEPDRS PIGDEHGEDEEKFTASINEEDKEKHAELLSLPPY GAEVFIGGLLPKDVIEEDLRDLCEPFGEIYEVKLIKN KDTGEFKGYAFVAFKTKEDAQKAIDELKNKEFKG KTVRCSLSDAKHRLFIGNIPKSWSEDQFKQMIHG VGPGAEMIDLKQDGLNPGRNRGFAFIEYYNNAC ADFAPQKMSNSSFKLEGNTPTVTWADPKSAPDS AAASQVKALYVKNIPESATPEQLKEIFQRHGEVT KVVLPPAKGGQGKGNFGFIHFGERSSALKAVKD TEKYEIDGQALEVTMAKPQTDKKSEAAASFAAGS NYLPQAAYGFAGNPYGSVGAGYGVAGGYQQPM IYGRGPMPSGMQMVPMVLPDGRIGYVLQQPGM QMASPVPRPRRNDRSNGSGGPQGRGNGSGNDR TRRFRPY
714	sp_065100_ujja	#N/A	#N/A	-0.859971	HEAT-INTOLERANT 1 (HIT1); CONTAINS InterPro DOMAIN/s: Vps53-like, N-terminal (InterPro:IPR007234); BEST Arabidopsis thaliana protein match is: Membrane trafficking VPS53 family protein (TAIR:AT1G50970.1); Has 898 Blast hits to 860 proteins in 251 species: Archae - 12; Bacteria - 82; Metazoa - 379; Fungi - 169; Plants - 68; Viruses - 0; Other Eukaryotes - 188 (source: NCBI BLink).	MSVPRSMDKSSALEYINQMFPTEASLSGVEPLM QKIHGEIRRVDASILGAVRQQSNSGTKAKEDLAA ATGAVEELMYKIREIKNKAEQSETMVQEICRDIKK LDFAKKHITTTITALHRLTMLVSAVEQLQVMASKR CYKEAAAQLEAVNQLCSHFEAYRDVPKITELREK FKNIKQILKSHVFSDFSSLGTGKETEETILLQQLS DACLVVDALEPSVREELVKIFCNRELTSYRQIFEG AELAKLDKTERRYAWIKRRLRTNEEIWKIFPSSW HVPYLLCIQFCKLTRTQLEEILDNLKEKPDVGTLLL ALQRTLEFEEELAEKFGGSSPAKEVNHDADNTE KGPNFSQAVSDIKKKYEKKLAANQGSGTDETESL KDFAVPGAGFNFRGIISSCFEPHLNVYVELEEKTL MEHLEKLVQEETWDAEEGSQTNILSSSMQVFLII RRSLKRCSALTKNQTLYNLFKVFQKILKAYATKLF ARLPKGGTGIVAAATGTDGQIKTSEKDERMICYIV NTAEYCHKTSGELAENIAKVIESVFAERVDMSDV QDEYSAVITKSLITLVQGLLETKFDAEMAAMTRVS WATLESVGDQSEYANGINTIVTSSIPVLGTLLSPI YFQFFLDKLASSLGPRFYQNIFKCKQISETGAQQ MLLDTQAVKTILLEVPSLAKQTSGAANYSKYVSR EMSKAEALLKVILSPLDSVADTYRALLPEGTTSEF QRILELKGLKKADQQTILDDFNKHGSSIKHPSMSP AVAIAPTPLAPPALSSPASSAITASREDVLARAAA LGRGAATTGFKRFLALTEAAKDRKDGPFRKLFNP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
715	sp_066580_wmid	-0.932104	-0.934846	-0.946313	Integrase-type DNA-binding superfamily protein; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: related to AP2 4 (TAIR:AT1G78080.1); Has 5953 Blast hits to 5757 proteins in 257 species: Archae - 0; Bacteria - 6; Metazoa - 0; Fungi - 0; Plants - 5926; Viruses - 6; Other Eukaryotes - 15 (source: NCBI BLink).	MEENYQTSSSFTTPIDTRDSVSALILTAGANNTNT LDSIFSYYPCQPSNSLTTTNNINSLLEPLGSSVYL KQRDLLQQFCLDNQGNKVGVYGSSSLSNQLFQN PSTTLNNPMKKKLYRGVRQRHWGKWVAEIRLP QNRMRVWLGTYETAETAAYAYDRAAYKLRGEYA RLNFPNLKDGATKLGFGDSARMNAVKSAVDAKI QAICQKLKKERANRIAKKKQKEATISESGGIKGGS GEEKSNKSTTHSGDSNLNENDSLMNSYSSSSSF CSTIVNDSGMVSPDISEEGWWSNGNGASPSSD SFTSYPMMTDDFGFEDCSLEKMPSFDPELIWEV LAN
716	sp_067140_fgiu	-0.927306	-0.883255	-0.933787	geminivirus rep interacting kinase 2 (GRIK2); CONTAINS InterPro DOMAIN/s: Calcium/calmodulin-dependent protein kinase kinase (InterPro:IPR020657), Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Calcium/calmodulin-dependent protein kinase-like (InterPro:IPR020636); BEST Arabidopsis thaliana protein match is: geminivirus rep interacting kinase 1 (TAIR:AT3G45240.2); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGCCGCFGFSRRHKAIARPKFRIDAFTSVLSQTP LLDDGDDDADADEDVDDGLYNGEDTSMTQGYD ESEVTGGAQRAEEIINFKIRNGLICRQFEVKETRK LVRSEDENGTKMINEYLREFKIGSGSYAKVVLYR SSIDGKHYAMKAFHKSHLLKQRVAPSETAMTDVF REVLIMKIVSHPNIVNLIEVINDPDTDHLYMVLEYV EGKWDCEGAGPPGGLGEAIMRQYLRDIVSGLMY LHSHNIVHMDIKPQNLLVSRNGTVKIADFSVSQVF EGDNDVLRRSPGTPVITAPECCLGLTYHGKAADT WAVGITLYCMITGKFPFIGETLQDTYDKIVNDPIDL PDNLDSQLRNLVEGLLCKDPDQRLTLDMVAEHP WVIGEEGPLPEFLCWCKQNSSGRDLHDVDKEVL ARGLEE
717	sp_067180_yuqe	-0.872422	-0.915319	-0.933315	proline-rich family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; Has 103183 Blast hits to 42587 proteins in 2006 species: Archae - 171; Bacteria - 19191; Metazoa - 36983; Fungi - 10058; Plants - 17697; Viruses - 3627; Other Eukaryotes - 15456 (source: NCBI BLink).	MEPRPPLMKLLILGLLFALFIVVESQQQSQQQPP QPPQQSQQSQNGQNPIDRQPPPQPPPQPPPTP TPPSSPFSPPPSSPTPRLTPPPSTLSPPPSTTIPRP PPSPLSPPPSSLSPPSSPISPPPSPFSPPPSPISP PPALPLASPPKLLPVGLSPPSPPPASPQPPPSPP LPPPPSPQPLPPQRSPRMTNPNTQANALRSNQR QNSKLNLGKKVGLLFAGIAVILQIGVVSFLAFKRK QLYKMRD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
718	sp_068730_zhiq	-0.892848	-0.951127	-0.889298	ACT-like protein tyrosine kinase family protein; FUNCTIONS IN: protein serine/threonine/tyrosine kinase activity, protein kinase activity; INVOLVED IN: protein amino acid phosphorylation, metabolic process; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, catalytic domain (InterPro:IPR000719), Amino acid-binding ACT (InterPro:IPR002912), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Serine/threonine protein kinase-like, ATMRK (InterPro:IPR015783), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: ACT-like protein tyrosine kinase family protein (TAIR:AT2G17700.1); Has 132714 Blast hits to 130570 proteins in 5062 species: Archae - 147; Bacteria - 14500; Metazoa - 50865; Fungi - 12053; Plants - 33999; Viruses - 512; Other Eukaryotes - 20638 (source: NCBI BLink).	MELIEGVGESSSPPRSFGGGGIGGSVLLDIRDDV YNRLIENGIEEVVSNPDFREILDVHFNKLPPSYGL DINIDRAEDVLLHRKLLALAKDPEKRPVFHVRCLE NFKSEDSEEHLALVNNHGERVQNNREEGREFEP CSKLEDLNLEVRQNSGMMKKSQEISSHVPTHEVI CSTPDKPKLLSQLSALLSDVGLNIREAHVFSTTD GYSLDVFVVDGWPVEDADSLRKAMEEAIDKSEG SWSGSSHSKFSVKKALAARGEPGNWEIDRRLLK IGDKVASGSCGDLYRGVYLGQDVAVKILRSEHLT EASEDEFAQEVAILREVRHENVVRFIGACTKSPD LCIVTEYMPGGSLYNYLHKNKRALKFSELLKFAID VSKGMEYLHHNDIIHRDLKSANLLMDSLQVVKVA DFGVARFLHQEGVMTAETGTYRWMAPEVINHQP YDQKADIFSFAIVLWELLTAKIPYENMTPLQAALG VRQGLRPEIPQEIHPKLCNLMQRCWDTAPSNRP PFSIIRVELEDLLENDQETSATVAGNLASGNIL
719	sp_069220_pcpn	#N/A	-0.904005	-0.88918	galacturonosyltransferase 6 (GAUT6); FUNCTIONS IN: polygalacturonate 4-alpha-galacturonosyltransferase activity, transferase activity, transferring glycosyl groups; INVOLVED IN: carbohydrate biosynthetic process; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 8 (InterPro:IPR002495); BEST Arabidopsis thaliana protein match is: los glycosyltransferase 5 (TAIR:AT2G30575.1); Has 1389 Blast hits to 1379 proteins in 238 species: Archae - 0; Bacteria - 400; Metazoa - 143; Fungi - 0; Plants - 830; Viruses - 0; Other Eukaryotes - 16 (source: NCBI BLink).	MKRFRRWQRIFILFLLSLSVFAPLVLVSIRLKTLTS EGKNKFVEDLPSIMQQYKSDVILLSAVGQEAREG VKEPRPEVYRDQNYSVVNFNQNQGGDSTVQGS RTSDVKVESGSERRGTQQLVGEEVKPIKPREEL VTVKTKDHSDQISSQHDHGVRAPRLTVHHSNNS RNETRKATDEKVKQMKDQVVRARAYLSFATTAG NTHLVKELKLRIKEVERAMGEARKDSELSRGRLQ KMKYMESTLSKASHVYPDCSNMVNKLRAMTHSA EEQVQSQRKQTTFLTELAGRTTPKGLHCLSMRL TTDYFMLQPEKQMLPNQQNVNNPDFHHYVTFSD NVLAAAVVVNSTVSNAMEPERVVFHVVTDSLNF PTISMWFLLNPPGKATIQVLNMQDFQWWNMFES SQEKQSSSDPRYSSLLNHLRFYLPEIFPQLGKIVF LDHDVVVQKDLSGLLTIDMKGKVNGAVETCQED KSSFRQMARFINFSDPLVARKFDANTCTWAFGM NLFDLRQWRHHDLTSSYRRFLLKGNGRPLWKA GSLPLGWITFYNRTLPLDQRWHLLGLGFDSVLRL EDIEQAAVIHYDGIMKPWLDIGIKKYKSYWNKYAN YDHPYMQQCNLHG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
720	sp_069380_grqa	#N/A	-0.0438	-0.074385	TUDOR-SN protein 1 (Tudor1); FUNCTIONS IN: RNA binding, nuclease activity, nucleic acid binding; INVOLVED IN: response to cadmium ion, protein secretion, response to stress; LOCATED IN: in 6 components; EXPRESSED IN: 12 plant structures; EXPRESSED DURING: M germinated pollen stage, seedling growth; CONTAINS InterPro DOMAIN/s: Staphylococcal nuclease (SNase-like) (InterPro:IPR006021), RNA-induced silencing complex, nuclease component Tudor-SN (InterPro:IPR016685), Staphylococcal nuclease (SNase-like), OB-fold (InterPro:IPR016071), Tudor subgroup (InterPro:IPR018351), Tudor domain (InterPro:IPR002999), Maternal tudor protein (InterPro:IPR008191); BEST Arabidopsis thaliana protein match is: TUDOR-SN protein 2 (TAIR:AT5G61780.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MAAAADRWLKGIVKAVPSGDTLVIMGIVKGDAIP PEKAVILKHIMAPRLARRDGDDEPFAWDSREYLR KLCIGKEIVFRADSNTSGKQNAPPGREFASVLLA DGSDVAKLAVQNGWAKVKTPSAQAKSEATSDLT ELLELEEKAKQQGLGRWSEVPGASEASIRNLPP SAMDPSNNFDAVAFASENKGRSLQAIVEQVRDG STVRVYLLPEFQYVQVFVAGIQSPSMGRMPPPE RTVEAEVSTDGSNGKTASEPRAFVTAAQKVASA AAPGVAVPPDPFGREGKHFTETRTLHRNVRIVLE QLDSFKNLVGSVYYDDGGEVKNLALELVKNGLA KYLEWSAKSLEAKDQKELKSAELEAKKQRLRMW TNYVPPATNSKAIHDQNFTGKVIEVVSGDCIVAD DAVPSGSPSAERRVNISSIKAPKLGNPRWNDNK GPESHSREAKEFLRQKLIGRQVNVSMEYSRKVP LGGATENAAAPSDFRVMDFGSVFLPAQSKGKGG DVTSVPSTAVAQEAGHNIAELLVSRGLATVVSHR DFEERSNYYDALLAAEAKARANKKGTHGAASPS NHIQDLTTASGKKAKDFLPFLQRSRRLSAVVEYV LSGHRFKLLIPKETCSIAFSFSGVRCPGRGEPYS EDAIAFMRRRIMQRDVEIEIETVDRTGTFLGSLWE SKTNMAAVLLEAGLAKFQKSFGTDRIADAHLLAQ AEESAKCQKLKIWENYVEGQEVTNGSSVPESRQ KEVLKVVVTEVLGGGKFYVQTVEDKMVVSIQEQL ASLEIGEAPIIGSFNPKKKGDLVLAQFSADNSWNR AMIVNVPRGGIVQSPKDEFEVFYIDYGNQEVLPY SRLRPLPPSVHSVPGSAQLCSLAFVKVPSLEEDY GQEAAEYLSELTLNSSRREFRAMIEEKDTSGGKSK GQGTGTVLLVTLVDVEASSSINASMLKVGLARLE KLKKWDSRERKSAIDNLEEFQEEAKKGRSGMW QYGDVQSDDEDSAPPMRKAAGKR
721	sp_069770_xshr	-0.871934	-0.817173	#N/A	uncharacterized protein	SRRPQPPPTGPPGHPTTEVIEPRSRNPQQNRVG GTVVADSQGRGTEHSSSVEHRGRGGAATSRAT EDSRSGGGEDSRRSSRRALLPSRSGGGEDRF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
722	sp_070970_tdng	#N/A	-0.6877767	#N/A	indole-3-acetic acid inducible 12 (IAA12); FUNCTIONS IN: transcription repressor activity, sequence-specific DNA binding transcription factor activity; INVOLVED IN: xylem and phloem pattern formation, response to auxin stimulus, embryonic pattern specification, root development; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aux/IAA-ARF-dimerisation (InterPro:IPR011525), AUX/IAA protein (InterPro:IPR003311); BEST Arabidopsis thaliana protein match is: auxin-induced protein 13 (TAIR:AT2G33310.3); Has 2001 Blast hits to 2000 proteins in 81 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 2000; Viruses - 0; Other Eukaryotes - 1 (source: NCBI BLink).	MFLSFVKRLRIMRTSEANGLAPRSEERSAKQRG RPIKSISSSCFIYREKKG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
723	sp_071150_aeqi	-0.924898	-0.980681	-0.939206	hua enhancer 2 (HEN2); CONTAINS InterPro DOMAIN/s: DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), DSH, C-terminal (InterPro:IPR012961), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), RNA helicase, ATP-dependent, SK12/DOB1 (InterPro:IPR016438), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: RNA helicase, ATP-dependent, SK12/DOB1 protein (TAIR:AT1G59760.1); Has 9841 Blast hits to 8173 proteins in 1240 species: Archae - 753; Bacteria - 3176; Metazoa - 1197; Fungi - 1159; Plants - 489; Viruses - 32; Other Eukaryotes - 3035 (source: NCBI BLink).	MAELPSTTTELPSTTTEAPSTMVENPPSSNAPVA EQQTLGKRKQSELEEPVPQQSSLEEQQKQQNQ KESQESSEKPGSKRRSSTRTCVHEVAVPKDFISD KDSSIHGTLANPSFNGSEAKKYAFTLDPFQQVSI ACLERNESVLVSAHTSAGKTAVAEYAIAMSFRDK QRVIYTSPLKALSNQKYRELSQEFTDVGLMTGDV TISPNASCLVMTTEILRGMLYRGSEVLKEVAWVIF DEIHYMKDRERGVVWEESIVFLPTAIKMVFLSAT MSNATEFAEWICNIHKQPCHVVYTDFRPTPLQHY VFPVGGSGLYLVVDENEQFREDNFLKLQDTFVK PGHGMKAAADKASGRIAKAGSGPKGGSDIYKIVK MIMERKFQPVIVFSFSRRECEQNAMSMSKLDFNT DEEKDIVEQVFKNAILCLNEEDRGLPAIELMLPLL QRGIAVHHSGLLPIIKELVELLFQEGLVKALFATET FAMGLNMPAKTVVFTAVKKFDGDSHRYIGSGEYI QMSGRAGRRGKDDRGICIIMVDEQMEMNSLRDM VLGKPAPLVSTFRLSYYSILNLMCRAEGQFTAEH VIRNSFHQFQHEKALPEIEKRIEKLEQEAATLDAS GESEVAEYHKLRLDLVLLEKKMMSEITRPERILYY LLPGRLIKIREGGTDWGWGVVVNVVKKAPAASG AVQRGGGYIVDTLLHCSPSTSDSGSRPKPCPPR PGEKGEMHVVPVQLPLIAALSKLRLSIPPDLRPVE ARQSILLAVQELEKRFPQGLPKLNPVKDMGIEDK EIVELVSQIEELEQKLFSHPMHKSQDFHQVKCFQ QKAEVNYEIQQLKTKMRESQLQKFRDELKNRSR VLKKLGHIDGDGVVQLKGRAASLIDTGDELLVTEL MFNGTFNDLDHHQVASLASCFVPCDKSNEQIQL RAELAKPLQQLQDSARRIAEIQRECRLDVNVEEY VESTARPYLMDVIYCWSKGANFSEVIQMTDIFEG SIIRLARRLDEFLNQLKAAASAVGEANLENKFGEA STSLRRGIMFANSLYL
724	sp_071240_odew	-0.883705	-0.909557	#N/A	uncharacterized protein	MLNIKHQGISSSDHYISRMLENGEANTPTISSTHL DSYGIPLLM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
725	sp_071300_xeww	-0.899708	-0.97236	-0.928033	MATERNAL EFFECT EMBRYO ARREST 58 (MEE58); FUNCTIONS IN: adenosylhomocysteinase activity, copper ion binding; INVOLVED IN: methylation-dependent chromatin silencing, one-carbon metabolic process, posttranscriptional gene silencing, embryo development ending in seed dormancy; LOCATED IN: plasma membrane, vacuole, membrane; EXPRESSED IN: 36 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: S-adenosyl-L-homocysteine hydrolase, conserved site (InterPro:IPR020082), S-adenosyl-L-homocysteine hydrolase (InterPro:IPR000043), S-adenosyl-L-homocysteine hydrolase, NAD binding (InterPro:IPR015878); BEST Arabidopsis thaliana protein match is: S-adenosyl-l-homocysteine (SAH) hydrolase 2 (TAIR:AT3G23810.1); Has 7964 Blast hits to 7958 proteins in 1477 species: Archae - 229; Bacteria - 2334; Metazoa - 639; Fungi - 145; Plants - 247; Viruses - 0; Other Eukaryotes - 4370 (source: NCBI BLink).	MATSLSVEKTTSGREYKVKDMSQADFGRLEIELA EVEMPGLMSCRTEFGPSQPFKGAKITGSLHMTIQ TAVLIETLTALGAEVRWCSCNIFSTQDHAAAAIAR DSAAVFAWKGETLQEYWWCTERALDWGAAGG PDLIVDDGGDATLLIHEGVKAELEFEKNGTIPDPA STDNVEFQLVLGLIKESLKVDPKRYHKMKERLVG VSEETTTGVKRLYQMQLNGTLLFPAINVNDSVTK SKFDNLYGCRHSLPDGLMRATDVMIAGKVGVVC GYGDVGKGCALALKAAGARVIVTEIDPICALQAL MEGFQILPLSDVVSEADIFVTTTGNKDIIMVDDMR KMKNNAIVCNIGHFDNEIDMAGIENYPGIKRITIKP QTDRFVFPETKTGVIILAEGRLMNLGCATGHPSF VMSCSFTNQVIAQLELWNEKDSGKYKKEVYVLP KHLDEKVAALHLNKLGAKLTKLTKDQADYISVPVE GPYKPAHYRY
726	sp_071360_riop	-0.912799	-0.971648	-0.911354	CONTAINS InterPro DOMAIN/s: Nin one binding (NOB1) Zn-ribbon like (InterPro:IPR014881), D-site 20S pre-rRNA nuclease (InterPro:IPR017117); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MDSQPPLSTPEAPPTVEPPVVCWSNIVKQTPSP TPPQPPKTPTKLFETCKSAKGIAVAIVDANAIIQG GERLPSLADKFVSVSEVINEIRDPTSRHRLSFLPF SVDTLEPSPESLKKVVSFARATGDLQTLSDVDLK LIALTYTLEAQVHGTSHLRDSPPPIHVLNVRRLPE KDLPGWGSNVPNLEEWEALDAATGGGPDHNSR ILPLKDLTLNIPVEDQISLQDGSVGNGTVDNDGAG LTTRRHRRYPPKKREVKLDGKMVADGVDASQG QHDDNDDGDWRPAVSRGTHRRFLRRKARRDM YEASLDDDEQQEVNKNTEEEVIMEAKCNGLTSD DRHVDQEISEDKKDTDGTNSAENISSILEQMRLA EDALKDHEEQDVQIGSHLEVASMDADLEDAHSG GEGDNDEVLDDLENGSVDTSYAEDDGSEQSWM LRSLSDSSVACITGDFAMQNVILQMGLRLLAPGG MQIRELHRWVLKCHACQNVTAEIGRIFCPKCGN GGTLRKVAVTVGENGVVLQARKQRFSLRGTKFS LPLPQGGRDAMSKNPILREDQLPQKFLYPKTKKK SNKDDDIFPGDSIFGHHTEKKTYNRPPVRQALAV FSGKRNPNDNHYSRAKH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
727	sp_071780_hjrg	-0.886386	-0.849572	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G48335.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MLENPSDLSSNPLSAPIKRYAPPNQRSRGLSRRK SGDRSERTSNLHLVDGDKNLGTGLGIDHSDTGS SKSVYESFRPRLTALNGCCRSEASQLLHERWNG AVHCYSDPTVDLSERPVMYTGSAASAWGNFRLP HQLMSAPQMDFLAELRQAINNSNASSMN
728	sp_071840_nwda	#N/A	#N/A	-0.869281	SPT4 homolog 2 (SPT42); FUNCTIONS IN: positive transcription elongation factor activity, zinc ion binding; INVOLVED IN: positive regulation of transcription, N-terminal protein myristoylation; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Transcription initiation Spt4 (InterPro:IPR009287), Transcription initiation Spt4-like (InterPro:IPR016046); BEST Arabidopsis thaliana protein match is: Transcription initiation Spt4-like protein (TAIR:AT5G08565.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGSVVAQIPTSFGHELRACLRCRLVKTYDQFRES GCENCPFFGLDKDHERVVECTTPNFTGIISVMDP SRSWAARWLRIGRFVPGCYTLAVSEALPEEFQAI CEDERVQYVPPKQK
729	sp_072230_dtom	-0.952658	-0.906443	-0.869781	plant U-box 9 (PUB9); FUNCTIONS IN: ubiquitin-protein ligase activity; INVOLVED IN: protein ubiquitination; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: U box domain (InterPro:IPR003613), Armadillo-like helical (InterPro:IPR011989), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: ARM repeat superfamily protein (TAIR:AT5G18320.1); Has 3234 Blast hits to 3002 proteins in 224 species: Archae - 0; Bacteria - 16; Metazoa - 312; Fungi - 206; Plants - 2395; Viruses - 3; Other Eukaryotes - 302 (source: NCBI BLink).	MAKTGVVSEGDSIRVLELKKELRLLVKTILEEEDH NYNNNNNDDYSTNSSCSSSVIDAIDKAKDLLSNL KEMKMKKTTSLRLDDGSFSSCPPHFCCPLSKQL MRDPVIISTGQTYDKSFIQKWLKAGHRVCPITQQ VLSHSLLTPNYLVREMISQWCESRGIENPDPVGY INEDGANEADRDYFSSLIDKMSKSIGEQKAAAKEL RLLTKRMPSFRALFGESTEAIPRLLKPLTQQRRG SFSSGSNELTHPDLQEDLITTLLNISIHDSNKKLVA ETPMVLKLLVDALQFGTPETRSNSAAALFTLSAL DSNKVLIGKSGALKPLIELLDEGHPMAVKDAASAL FNLCIIQENKGRAVRDGAVRVIFKKIKNGVHIDELL AILAMLSTHQRAVEELSDLGAVPCLLANLRESSC ARNKENCIAILYTICVNDRTKLREMRDEERAYGTL SQLAQDGTSRAKRKASGILDRLNRAHNLTHTA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
730	sp_072290_ehhq	-0.930707	-0.893728	-0.920917	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: ribosome biogenesis; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: AARP2CN (InterPro:IPR012948), Protein of unknown function DUF663 (InterPro:IPR007034); BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT1G06720.1); Has 2741 Blast hits to 2088 proteins in 291 species: Archae - 2; Bacteria - 131; Metazoa - 833; Fungi - 650; Plants - 171; Viruses - 49; Other Eukaryotes - 905 (source: NCBI BLink).	MGGSRAQVNKAHKTRFASKSSRNIHKTSLKDKH QNAKSDRFGKGARAARIQRNNMIRDQKRTAVLN EKRASSGSASPPLVIVLFGLSANVDLNAVERDLL GLLSSECGSEASPTVVSSEYKIRATVLKAPHGDL SPCMDMAKVADLMAFVTSASSPCEDGSSDYIDS FGLQCLSVFKAIGLPSTVVLIRDLPSDLKGRNESK KFCNSSLASEFPDPCKFYPADTKDDLHKFLYLLK DQRLKSPHWRNQRAYVASPKLEMLDHDSVSGK CTLCVTGYIRGRNLSVNQLVHVSGAGDYQLSKIDI LRDPYALNPRKEQDTMDADDAPDLQVISSLVPDD SKQEPLHVENVPDPLAGEQTWPSKKEMEDAEK NLKHKEVKKRRLPPGTSDYQAAWIVDESDDDDD VGSVDESNDGMALDDEEDGFASKEISKSFDLDE DQGSLDLGESDGETETDSEMIEGPTREQIEHDIK RLKDAHAEDEEYPDEVDTPLDVPARKRFARYRG LKSFRTSSWDPKESLPVEYSRIFAFDNFTRTRKH VLAKALEWDNKSKEETVSPGTFARLYIKDVPQAV ASTLYSRAKMIPIIVCGLLQHESKMSVLHFSIKKH DSYHEPIKGKEELIFHVGFRQFVARPIFSSDNFNS DKHKMEKFLHAGRFSVASIYAPISFPPLCLVAMKK LGDMGSPVLAAVGSLISVDPDRILLKKIILTGYPQR VSKVKATVRYMFHSPEDVRWFKPVEIWTKCGRR GRVKEPVGTHGAMKCLFNGVLQQSDTVCMSLF KRAYPKWPEHKFPLSNF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
731	sp_072340_uoit	#N/A	-0.831111	-0.877558	trehalose-phosphatase/synthase 9 (TPS9); FUNCTIONS IN: transferase activity, transferring glycosyl groups, trehalose-phosphatase activity; INVOLVED IN: trehalose biosynthetic process, metabolic process; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: HAD-superfamily hydrolase, subfamily IIB (InterPro:IPR006379), Glycosyl transferase, family 20 (InterPro:IPR001830), Trehalose-phosphatase (InterPro:IPR003337); BEST Arabidopsis thaliana protein match is: trehalose-6-phosphatase synthase S8 (TAIR:AT1G70290.1); Has 4900 Blast hits to 4828 proteins in 1062 species: Archae - 56; Bacteria - 2863; Metazoa - 236; Fungi - 655; Plants - 621; Viruses - 0; Other Eukaryotes - 469 (source: NCBI BLink).	MGSKSYGNLLGLVSGDLLDIPCTPRTLPRVMTVP GIISDLDGRESSDEDVDASIVRERTIVVANMLPLH ASRDTETGKWRFTVDEDSILLQLKNGLTPNTEVI YIGSLKADIDINEQDEVSQKLLREFKCVPTFLPQD LQKKFYHGYCKQYLWPLFHYMLPMCPDHGDRF DRALWQAYVCANKIFADKVMEVINPEDDYYWVH DYHLMVLPTFLRKRFFRVKLGFFLHSPFPSSEIYR TLPVRDEILRGLLNCDLVGFHTFDYARHFLSCCS RMLGLDYASKRGHIGLDYFGRTVFIKILPVGIHVG RIESALNLPCTSVRVKEIQEQFKGKKVIVGIDDMDI FKGISLKLLAIEQLLLQNPEFQGTLVLIQIINPARSS GKDVQEAKRETYLIARRINEMYGSPDYEPVVIIDR PVPRFEKSAYYAVAECCLVNAVRDGMNLVPYKYI VCRQGTPTLDEALGVKSDTPQTSMLVVSEFVGC SPSLSGAIRVNPWDIDAVAEALNLALTMPDAEKH LRHEKHYRYVSSHDVAYWARSFVQDLERACQN HYNKRCWGFGLGLSFRVVSLSPSFRKLNVDYIAS AYRRTSKRVIFLDYDGTIVPQTSMNKCPSPEVFS VLNALCNDPKNTVFVVSGRGKTQLDEWLSPCQL LGIAAEHGYFLRPNKATEWESNTAELDWKEIVEP IMRSYTEATDGSNIETKESALVWHHQDADPDFG SWQAMELMDHLENVLANEPTVVARGHNIVEVKP QGVSKGLVVEKVITTLINGGKRPDFLVCIGDDRS DEDMFESLLKFSSPTSSGLPEVFCCTVGQKPSK AKYYLDDTSDVLRLLAGLAAAASTQKPRYNPHAQ VSFESTF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
732	sp_072350_yyue	#N/A	#N/A	-0.87713864	K+ uptake permease 6 (KUP6); FUNCTIONS IN: potassium ion transmembrane transporter activity; INVOLVED IN: potassium ion transport; LOCATED IN: plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Potassium uptake protein, Kup (InterPro:IPR018519); K+ potassium transporter (InterPro:IPR003855); BEST Arabidopsis thaliana protein match is: Potassium transporter family protein (TAIR:AT5G14880.1); Has 3423 Blast hits to 3365 proteins in 1017 species: Archae - 13; Bacteria - 2376; Metazoa - 1; Fungi - 98; Plants - 811; Viruses - 4; Other Eukaryotes - 120 (source: NCBI BLink).	MDPESGLSQNQQVKKKESWRQVLILAYQSLGVV YGDLSTSPLYVYKSAFAEDIQHSESNEEIFGVLSF VFWTLTIVPLLKYVFIVLRADDNGEGGTFALYSLL CRHARVNSLPNCQLADEDLSEYKKDGVVVPAQT SFGSRLKSILEKHRVLQKLLILALIGACMVIGDGV LTPALSVFSAVSGLELSMSHEHHKYVELPVTCVIL IGLFALQHYGTHRVGFLFAPIVVTWLFCISTIGLYN IIHWNPHVYKALSPVYMYKFLKKTQTGGWMSLG GILLCITGSEAMFADLGHFSQSSIKIAFSFVVYPSL ILAYMGQAAYLSRHHVITTTYQIGFYVSVPEKLRW PVLGIAILAAVVGSQAIITGTFSIIKQCQALGCFPR VKIVHTSSKIHGQIYIPEINWILMILCLAVTIGFRDT KRMGNASGLAVITVMLVTTCLMSLVIVICWRKSV FVALALAFVLFFGAFEALYFSASLIKFLEGAWVPIALS LFFMLIMYVWHYGTLKKYEFDVQNKVSINWLLSL GPSLGIVRVRGLGLIHTELVSGIPAIFSHFVTNLPA FHQVLVFLCVKSVPVPHVRPEERFLVGHIGPKEY RLYRGIVRYGYRDFHKDDFEFENDLVCSIAEFIRG ESNKMNGLKDESDKDTDEKMTVVGSPSTHVDGI RMHEDVEVEDIPETSEMKEIRSPLTTPKKRVRFLL PETPKMEDKTMEELQDLMEAREAGIAYILGHSHV RAKQGSGWIKKLAINIGYDFLRRNCRAPSYPLSV PHASTLEVGMICHM	
733	sp_072400_ijms	#N/A	-0.888838	-0.888093	lipase class 3 family protein; FUNCTIONS IN: triglyceride lipase activity; INVOLVED IN: lipid metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Lipase, class 3 (InterPro:IPR002921); Has 393 Blast hits to 389 proteins in 69 species: Archae - 0; Bacteria - 12; Metazoa - 12; Fungi - 52; Plants - 275; Viruses - 0; Other Eukaryotes - 42 (source: NCBI BLink).	MDRYLLAEAGDTLFASVVGTKQYKDLVANANIFQ GALFDDDDMEIMEETEIVESDIADGRKGYLYSHA KTETKPKQLKHKLRPAAHRGFMARAKGIPALELY RLAQKKKCKLVLCGHSLGGAVKYVC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
734	sp_073500_mwyk	-0.925561	-0.948369	-0.965091	SET domain protein 25 (SDG25); CONTAINS InterPro DOMAIN/s: SET domain (InterPro:IPR001214), GYF (InterPro:IPR003169); BEST Arabidopsis thaliana protein match is: homologue of trithorax (TAIR:AT2G31650.1); Has 5838 Blast hits to 5683 proteins in 501 species: Archae - 3; Bacteria - 461; Metazoa - 2434; Fungi - 507; Plants - 1016; Viruses - 2; Other Eukaryotes - 1415 (source: NCBI BLink).	MVSSTHLFDTIKHRHKPEVDLLSAKRLKVSASGSI DSLSCLETDEFDEKTDCVTSSSPACLSSSCDSEG QIPSSSSLERSCQVNGHADAVSLASGAAGISYPR KNDNEHGSSDFVNGWMYLNEQGQMCGPYIQQ QLFEGLSTGFLPDELPVYPILNGALLNPVQLKFFK QYPDQVTTGFAYLRSTTPNVSITTSSHIAPGQDT HKDTGSVLLNYDTQSHVQQASHSTDAIVSDSKLS PRTAVVQQQSLSWEEKCWLFLDGYGRVTGPHS LSELYYWHQYGYLHSSVMIRHTDNKCGPFSLLSL INACGKESSSTANTSEKESISSSELISGVSEEIGS QLHDGIMKAAHRYVLDEIIGNVIADYVAMKKAQK QLKNEHGREDLIFERNKTKDGENNSSASHAAITS CHDDASTSATAKETLEEPSSFQHMKSIGSVENY QGATEAISRVLFDYCMQVMWNAVFYEPVANCTL TWRKNKRWSHNSTEEAVALDKNSDFQKIVVVEV EASEPGVPADELDCPPGFHSIVQDSLLCDSSLIA QSHFREDKLSSPSMQHQDEDREGTIGRVEEELF MLAKTSMDDFLKISIEKEVFKLNKCLKDNGTHEVS VENSGQLSKDNDSGYAASEMLQGLVSRRICLDE SHIISQGTGLLHQPAVCPHKDSSSDFLGTGFRSL GAPAITAIHCCKTDEPLPPGIADLTIDPSAQDLQKF LLDHLKIPIQAIRSHEAADLAHGNSSSNFIGHAFKR LGMPICSVIDDHRTDEAQLPGAGGEVKDIVSPTIS EFRPLHLGEGIPKQKKFMGFALCRQKLHDDVLRE WLSSFHDELHCLLKQRDLVQEELIIRTEECSKLTH PMQPELDHLVERSKSCYSSGPSEVSVVEEFTYS RRKKVLKKSGAEPSSPPLRDFKLHHQSSKKFGK PIISNNVSCSGALGIKNVKQKAKGVCQRETGLSS LTSKLGAVKRSLGKDRQSIRSSNGRRLQSFALPS QSTVAVDDPKHGVEKVLDMEAESNDIEELSSSSS IHSNAKKDKLANDHSQTVKNSSTVSKLKRKNLVG ESSSLPSKAPKLVDAGAKQAEVKVAERRTKLKKS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
735	sp_073920_ncij	#N/A	-0.866618	-0.863409	Chaperone DnaJ-domain superfamily protein; FUNCTIONS IN: heat shock protein binding; INVOLVED IN: biological_process unknown; LOCATED IN: cytosol, plasma membrane; EXPRESSED IN: male gametophyte, guard cell, cultured cell, pollen tube; EXPRESSED DURING: M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Heat shock protein DnaJ, N-terminal (InterPro:IPR001623); BEST Arabidopsis thaliana protein match is: Chaperone DnaJ-domain superfamily protein (TAIR:AT4G12780.1); Has 79558 Blast hits to 47178 proteins in 2717 species: Archae - 253; Bacteria - 22646; Metazoa - 24220; Fungi - 7704; Plants - 5305; Viruses - 398; Other Eukaryotes - 19032 (source: NCBI BLink).	MNDFQGLLSTDYGFKPQGKAAPMAASKSGTNS SNSAFNFEIGGSKTASFPSNSKNPTSSSLFDDHD TLFRSNRKTTSSDDFGILDDVFATGGIKSNSAKSS NNNSVNMDSIFGGGGGGGGESGPKFSSLPVFD KPVYDDEDSSSAKFENIFAPPSSAPAKKTDAFDD LLGGLGKKEQQSTRSAVGPGSKGADKGFSSFDD LLPGFGGGSTSSTSRNPSDASKKTSSVIDDPFVG LDSTSPPVSSSGIFTDPLEEIGNLNNSSRKTAESS SSGTAFDDIDPLGGFGKSVPSFSAEINGEEKEAM RTRVSGRQTSVSGDASRKSSVRSPDRGSEPKFP VEDYSEPSVFEMPSVSAETSRSVPNASFPTHSN ASFRESNSHGHSSPNSEENMGSSDDIWLTVSEV PLFTHPTAAPPPSRPPPPRPTHFSASYPASNVHK KVNDYPSFPGPGPSFHSPKSAPVEARGSFASPL DELEEFAMGRNPSNEDATHASSEGGSGEDFDS NSVAAASAAAMKEAMDRAEIKFRHARGVREREN AKASKSRESVQVEKEENPTYDAPEREFREHQER LERERQQREMEEEERRRTKEEERRRAEKEWEE KEREKRKVEKERQFGRQAVEWAAKEARERAAIE ARARADRVAVQRAQAEARERAEKAAVHRAQSE ARERAAAGAKERAEKAAAEAREKAKAEAREKEA REREAREKEAREKAATAAKANQSKNENDLESFF SMGRASSAPRPRAPSSDPFEAQFQTKPAAEATR ASVNSSSSMRKASSSTNIVDDLSSIFGGPPSSGG FQEVDGETDDRRRARMDRHQRTQERAAKALAE KNARDLEVQREQAEKNRISETLDVEIRRWAAGK EGNLRALLSTMQYVLWPECGWQPVSLTDLITGA AVKKAYRKATLCIHPDKVQQKGATLQQKFVAEKV FDLLKEAWNKFNSEELF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
736	sp_075180_rzyq	-0.885853	#N/A	#N/A	DEAD box RNA helicase family protein; FUNCTIONS IN: helicase activity, ATP-dependent helicase activity, ATP binding, nucleic acid binding; LOCATED IN: cellular_component unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), RNA helicase, ATP-dependent, DEAD-box, conserved site (InterPro:IPR000629), WW/Rsp5/WWP (InterPro:IPR001202), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEAD box RNA helicase 1 (TAIR:AT3G01540.4).	MAATATVSPGPRYAPEDPSLPKPWRGLVDGKT GYLYFWNPETNVTQYERPGSSSHGCSIPSRPFS GYVSSSAQVQQSSNGHKDDNDDKYNRGSSGER IQQSARGGNTPNEAVSERSTIEHVPVDAVSGLSA EAYRRRHEISVSGTGVPPPFITFEATGFPSELLKE VHQAGFSAPTPIQAQSWPIAMQGQDIVAIAKTGS GKTLGYLLPGFMHLKQRRNNPQMGPTVLVLSPT RELATQIDEAVKFGRSSRIACTCLYGGAPKGPQ LKELDRGVDIVVATPGRLNDILETRRVSLNQVSYL VLDEADRMLDMGFEPQIRKIVKELPTRRQTLMYT ATWPKEVRKIAADLLSNAVQVNIGNADELVANKSI TQHVEVLSSMEKQRRLEQILRAEEPGSKIIIFCST KKMCDLLARNLSRQFGAAAIHGDKSQSERDHVL NQFRSGRSPVLVATDVAARGLDVKDIRVVVNYDF PTGIEDYVHRIGRTGRAGATGVAYSFFNDQDSKY ASDLVKILEGANQKVPQEVRDMALRGGGGGMG RFRRFGSGSGGRDGGRGGRSDSSYGGRDGGR GRDRYDSYDRGYQAARGRSRSPSPTGQKAWGG NQVKDQGWSRGRSRSRSPEKINKGQPGKSFHE MMMEKTRSSPMFPQKTPAADYGNGQSAWGKS NDEEEEEGAIPS	
737	sp_075680_pzyf	#N/A	-0.883059	-0.853416	ser/arg-rich protein kinase 4 (SRPK4); FUNCTIONS IN: protein kinase activity, kinase activity; INVOLVED IN: response to oxidative stress, protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase-like domain (InterPro:IPR011009), Protein kinase, catalytic domain (InterPro:IPR000719), Tyrosine-protein kinase, catalytic domain (InterPro:IPR020635); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT3G44850.1); Has 38460 Blast hits to 29477 proteins in 1034 species: Archae - 2; Bacteria - 1392; Metazoa - 15648; Fungi - 7270; Plants - 6510; Viruses - 16; Other Eukaryotes - 7622 (source: NCBI BLink).	MDNRRKETVSVTADTEELREKGEPREKEAAVEEE EEEEEVSESGSDYTSEDEGKDDYRRGGYHAVRI GDSFKNGAYVVQSKLGWGHFSTVWLAWDTNHS RYVALKVQKSAQHYTEAAMDEITILKQIADGDPD DKKCVVKLLDHFKHSGPNGQHVCMVFEYLGDNL LTLIKYSDYRGMPMDKVKEICYHILVGLDYLHKQL SIIHTDLKPENILLLSVIDSSKDPRKTGAPLVLPNS KDKTLYESSSTKESKVSNGDLTKNQKKKIRRKAK RAAQGCAGKEAAADSEQDPETSDVDGVKRTSG GSSVHKRGSRSTRRKLLESVDLKCKLVDFGNAC WTYKQFTNDIQTRQYRCPEVILGSKYSTYADLW SFACICFELATGDVLFDPHSGDNYDRDEDHLALM MELMGMMPRKIALGGRYSRDFFNRYGDLRHIRR LRFWPLNKVLVEKYEFNEKDANDLAEFLIPLDFV PDKRPSAGECLSHPWIAGGLTGLKASPPESQEV DVTITDKKKKKEKDEREAMEKGMGSIAITIDSKPSK QPLSSPKPSKAAVIGSSR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
738	sp_075870_pinw	#N/A	-0.8671475	#N/A	VACUOLAR MORPHOLOGY 3 (VAM3); FUNCTIONS IN: SNAP receptor activity; INVOLVED IN: in 7 processes; LOCATED IN: late endosome, trans-Golgi network transport vesicle, plant-type vacuole membrane, vacuole, plant-type vacuole; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Target SNARE coiled-coil domain (InterPro:IPR000727), t-SNARE (InterPro:IPR010989), Syntaxin/epimorphin, conserved site (InterPro:IPR006012), Syntaxin, N-terminal (InterPro:IPR006011); BEST Arabidopsis thaliana protein match is: syntaxin of plants 23 (TAIR:AT4G17730.2); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSFQDLEAGRRPLASRRDLINAKQDPTQAVASGI FQINTAVSSFQRLVNTLGTPKDTLQLREKLHKTR LHIGQLVKDTSDKLKQASETDHSNNVSANKKIAD AKLAKDFQAVLKEFQKAQRLAAERETAYTPFIPQ AVLPSSYTASEIDVSSDKTPEQRALLVEARRQDVI LLDNEITFNEAVIEEREQGIQEIQHQIGEVNEIFKD LAVLVHDQGAMIDDIGTNVENAQAATVQAKSHLA QASKTQRSNSSLTCLLLVIFAIVLLIIIAA
739	sp_076260_dyus	-0.8913952	-0.8447755	-0.8852274	LRR and NB-ARC domains-containing disease resistance protein; FUNCTIONS IN: ATP binding; INVOLVED IN: apoptosis, defense response; CONTAINS InterPro DOMAIN/s: NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: NB-ARC domain-containing disease resistance protein (TAIR:AT3G14470.1); Has 38623 Blast hits to 22788 proteins in 858 species: Archae - 54; Bacteria - 2838; Metazoa - 8237; Fungi - 658; Plants - 25379; Viruses - 2; Other Eukaryotes - 1455 (source: NCBI BLink).	MDIVGTLLSAAQTLLAALQCSELQEILSIFGYKSQ LDDLQRTVSAINAVLLDAEAKQELSHEAKHWLEE LKGAVFEADDLLDEFVTLAEQKQLMEAGGSLSKK WPKNTTNVVNLDIETEGLYLRNKEHLQGLKFDFL HRKVDGDVDTKEARRLMEELQPHSKLKELTVRR YSGVRMPSWTTLLPNLVFLALQNCEELGYLPCL GNLRHLKHLKLEGLEKLEYIGISPSVLGSAEGVPF FASLEYLELSCLPKLKGWAGVGAGDDLLFLDDC ENIQLQLRLCLSQLKVLLIEVCPELTGMPPCPGLK YLQLRKFNRRLRIINTSSMFCHDISCPSPSRSLFD PENSRSNRQWTVDIDDVAWLNSLPIEAFQGLER LAIKGQLDEAIGEAVESLEEVKEVFSNCSSSLQDL DIIGFDKLRSMPGALEHLTALKKLTIWNCPNVRLS EEREDGTPWRSLHHSLISLGFVNLPQLVNLPIWM QSLAALEKLHIEDCKTLESMPSWMSKLTSLSTLR LLECSDSLKKRCQKDPPGEDWPYIQHIKEIDVRE DRC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
740	sp_076300_hztg	#N/A	-0.870166	#N/A	RPM1 interacting protein 2 (RIN2); CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841), Zinc finger, C3HC4 RING-type (InterPro:IPR018957), Ubiquitin system component Cue (InterPro:IPR003892); BEST Arabidopsis thaliana protein match is: RPM1 interacting protein 3 (TAIR:AT5G51450.1); Has 2005 Blast hits to 1973 proteins in 236 species: Archae - 0; Bacteria - 2; Metazoa - 699; Fungi - 246; Plants - 674; Viruses - 9; Other Eukaryotes - 375 (source: NCBI BLink).	MGVRYLAISAVCTVVSMIGLQCWTEISLDKFKAD GLIGNDTLYSENASRIIELLLLRSHATITLAAICVINA FVLFILSIKTIFFGELYAPETRKLVERLVNYVIYKGT FLPLIIPPTIFQSSLWLTWLTVLCFLKMFQALARD RLERLNASPSATPWTYFRVYSALLLVLAFDSLWI WFCAAVTGPQVSSTYWLLFFEPLSIAFETLQAIV VHGFQLMDVWLHDSAAETSDCQRSKLFDTTAG SLWEWKSSVVRNLGFFLDMMTLLLMALGHYVLIW WLHGMTFHLVDAILFLNIRALLSAVVKRIKGFIKLK IALGALHEALPDATSEEIQAYDDECAICREPMAKA KRLSCNHLFHLACLRSWLDQGLNEVYSCPTCRK PLFVGRSESEASSHSRAVSSDEQLARQLSSGVD RVNAEGGGVFPNQNQTPPEVNAWRGAGADSS WLQPWPSQGLDVAGPSTAIRSAGLGRVQMMMR HLASVGENYAQTALEDTSWSLWPTSSSQATPSN TTFPPAAHRYPATTTTNTAHRRTAPRPANDSIAN LLAMAETVREVLPHVPDEMIFQDLQQTNSVAVTV NNLLQM
741	sp_076350_znwf	-0.909867	-0.861755	-0.883481	LUC7 N_terminus domain-containing protein; FUNCTIONS IN: RNA binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: LUC7 related (InterPro:IPR004882); BEST Arabidopsis thaliana protein match is: LUC7 related protein (TAIR:AT3G03340.1).	MDAQRALLDELMGAARNLTEEERKGYKEIKWDD KEVCAPFMIRFCPHDLFVNTRSDLGPCPRIHDQK LKESFENSPRHDSYVPRFEAELVQFCEKLVADLD RKVRRGRERLDQEVEPPPPPPISAEKAEQLSVLE DKIKNLLEQVESLGEAGKVDEAEALMRKVDSLNV EKAILTQQPLNPATMLTQEKKMALGEICGSFLVA NDAMERTQSHITGKQHIGYGLVRDFIAEFKAAKE KAREEERLAREKEAEERRKQKENENEGKNRRSI SSERDCHRDRDRDYSRDRDRSRERNNRGNRDE GRGMDRRRNNDHRNGRDGGGRNRYYGRDRSR SRSPVRHGHRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
742	sp_076570_mgcf	#N/A	-0.884383	#N/A	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein; FUNCTIONS IN: asparaginase activity, hydrolase activity; INVOLVED IN: glycoprotein catabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase T2, asparaginase 2 (InterPro:IPR000246); BEST Arabidopsis thaliana protein match is: N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein (TAIR:AT3G16150.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MAKIEMTFVGFPAGFQSLLIIIIITLLVANTVVANEV EENNGKYYPIVVSTWPFVDAVRAAWSVVDGGFS AVDAVVQGCSACEQLRCDGTVGPGGSPDENGE TTLDALVMNGATMEIGAVAAAMRYVKDGIRAARLV MQYTEHTLLVGEQASVFAISMGLPGPTNLSSEES LHKWNQWKDNNCQPNFRKNVFPSDSCGPYHPK ENVGFGETSCPTAYLVETMTPKVSPINIHNHDTIS MAVIDKTGHVAVGTSTNGATFKIPGRVGDGPIPG SSAYADDEVGACGATGDGDIMMRFLPCYQVVES MRRGMEPKFAAEDAISRILKKFPDFVGAVFADNS SCIGRLCHAESAAEIKFSDVYYATELADWGAVRQL TKCCFLHQDSEMYLFVVHAMQRTKTKASFWHLV EYTFGHRDQLVFVNPKSGKGNGQNIWGVVAPIF SRAKVRAQVTVTERAGHAFDLVSSMSNRELISYD GIVAVGGDGLFNEILNGLLSSRFKTEFPPVPPEF MASPGEDPSTSVPIPNEINAADEDVEFSFPHEWF RFGLIPAGSTDAIVCTTGARDPITSALHIVLGKRIG LDIAQVVKWKATMDSDFEPSVHYTASFAGDNTE LRDLRRF	
743	sp_077240_nted	-0.940892	-0.883598	-0.931261	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G23150.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MQVLSDVKTFQVLLSEICLEQHHRDYRIAGCSEL LTFRLCQSRPKVLEKSSGLVAGVALTGHSWHWE KSLH
744	sp_077350_qfhz	-0.95085	-0.903914	-0.921467	AT-hook motif nuclear-localized protein 1 (AHL1); FUNCTIONS IN: DNA binding; LOCATED IN: mitochondrion, nucleolus, nucleus, cytoplasm; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF296 (InterPro:IPR005175), AT hook, DNA-binding motif (InterPro:IPR017956); BEST Arabidopsis thaliana protein match is: AT hook motif DNA-binding family protein (TAIR:AT4G22770.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MATATPPVSNSMAGVGERVPTFTSSDKKKRGRP RKYGPDGAAAAAAALSPMPISASIPLSGDFSGWK QPRARPFQNIKKAHKVESVQTESLGDRLAYFVG ANFTPHVITVNAGEDIAMKVLSFSQQESRAICVLS ASGSISNVTLRQPNCSGGTLTYEGRFEILSLSGS FMPTDNAIMKSRSGGMSVSLAGPDGRVLGGGLA GLLVAASPVQVVLGSFLPGHHQEQKPKKQRVVE PVTPAMPPPIVNNHQVPDGGGYGGVHNKVNFT SPPPPSFHTENLPSMNSIHLDPMDLEADHNNVST PEGDSDDFDDDDDDSLQI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
745	sp_077630_mkip	#N/A	#N/A	0.8619023	SPLAYED (SYD); CONTAINS InterPro DOMAIN/s: DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase/SANT-associated, DNA binding (InterPro:IPR014012), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021), SNF2-related (InterPro:IPR000330); BEST Arabidopsis thaliana protein match is: Homeotic gene regulator (TAIR:AT3G06010.1); Has 37993 Blast hits to 29119 proteins in 2440 species: Archae - 313; Bacteria - 8300; Metazoa - 11402; Fungi - 6833; Plants - 2357; Viruses - 252; Other Eukaryotes - 8536 (source: NCBI BLink).	MYKELNDVADVPTASDAVDLRSGSTLEAQSRED EQIHGSDDGKVDAFDISTASGAVDIGIVSTLESHS RVDEQIHGPDDGKVDAVDAPTASDAVGLGIASTL ESHSREDGQYHGLGDGKVDAVGAPTASDAVSL GIAATLESHSREDGQFHEPGDGKVDAVDAPTVS DDVRLGIASTLDSHSNEDEQFHGHDVGKVDAAD SCAVAAPSLVLEMEVSRDTNDDTADTNVGGPDS DAAATSSLESKQKEKLHGPDHDSGELEIGDAPTV CCVLGSEAAAISTTKLNEKEEFLGSEVGYTVNDA ADAPAVSLAVVSTTVTSTWGLEDKGELCSPSDIN AVHDAGNAPVVLSSVISDIAVTPMLASIVTEEPRV ADNSNGDEHVAGPAGNAPSTPNIVEFDTAVCSIQ QSKAIKECHEPVDGYTEIDVVGSVTVVTSPFETK EKEELEGHNDGGAELDVGESPTDSTIFGSDIAAV LTSVLKTKEDCLGPEAGIVEHDAVDVSSACIVGF DKQAASSSKDKKECHGPDSYDKLKADDTYTNLH VGSDLAETSTLIEEECNGLDRGNGELDAVDAPTA FHAVRSGASSSESKKECHGPKDGDVKHDDTGDI FKGSDTPETSTLSSTDKEKCHGCTD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
746	sp_077930_opdp	#N/A	#N/A	-0.9934583	ATP binding microtubule motor family protein; FUNCTIONS IN: microtubule motor activity, ATP binding; INVOLVED IN: microtubule-based movement; LOCATED IN: chloroplast; CONTAINS InterPro DOMAIN/s: Kinesin, motor region, conserved site (InterPro:IPR019821), Kinesin, motor domain (InterPro:IPR001752); BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT5G60930.1); Has 10625 Blast hits to 10139 proteins in 419 species: Archae - 2; Bacteria - 255; Metazoa - 4769; Fungi - 1373; Plants - 1690; Viruses - 3; Other Eukaryotes - 2533 (source: NCBI BLink).	MEWNSPEPPRSITVRRNPHRSARPTTTNAAPKT STSHSLSSKNLKIPNFPLDDILSMDIQLNTNPKIPS PNPPNSLNVFLRVRPLPIAPPKSKQIKPSKPSKSC LTINDSQSITISPPIRDNRLKSEVFNGFSHVFSSDS SQDEVYEKLMRPMVEEFLNGRSGLLAALGPSGS GKTHTVFGSAKEPGLVPRALRRIFESDAGKSSLV SRSIYLSVFEIHSEKGKGEKLSDLTPDGSDIYMQQ SAVKGLKEILVKDVAQAESLLTHAKSKRATATTN SNSQSSRSQCIINRCAAEDTEGESEAVLTFVDLA GAEREKKTGNKGGRLLESNFINNTSMVFGLCLR SLLEHQKNPKKPLQKHFKSSLLTRYLRDYLEGRK RMALILTVRSGEEDYFDTSSLLRQASPYMEIKFC DIEETLNLPRNKRRHQIICNEQKKRVKLTIADGRN DVTVISRDSETFQTKEAKSREILDGNNGIPVSRVD ENLLPEVNCTDSNIHRQYYFMQGFSKALWVVLK QYKERVKTMQKDINILTENLRVGDAKYLVMEEEL LKLKSQCSCPNQVIGVTSCSMGGLNSEPSFCEE EFTDHGKVVSTDVQGLACLECKEKEILHSSGEKC TLHVCELFGVAAPEEPHECSVSECKTTNLSCEKA LHTEGKVDVTLIPKSSENHDLEEESQSGSVTDGL QELQGSRCVNALKVHLGDSSSGPKSILSVDESID TDFILTSPHKANLSSNELKIFSPHNHQSLMEYQTP HDMSKIDCHESGQRKKCDIQSTDAAACLHRSHV VDLIDQGESCGRDTPNNIDKENAGNGPGSKSRL EKKETSHSKPSKAGRPKRRLMPASSLLLRDTGSI GAIEENIHFRGGKGGQKLVEDEHKRTKGNACLM RLLKNHLPR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
747	sp_077950_eczu	-0.8837445	#N/A	#N/A	MATERNAL EFFECT EMBRYO ARREST 5 (MEE5); FUNCTIONS IN: translation factor activity, nucleic acid binding, GTP binding, translation elongation factor activity, GTPase activity; INVOLVED IN: regulation of embryo sac egg cell differentiation, embryo development ending in seed dormancy; LOCATED IN: nuclear speck, cell wall, plasma membrane; EXPRESSED IN: whole plant, cultured cell; CONTAINS InterPro DOMAIN/s: Protein synthesis factor, GTP-binding (InterPro:IPR000795), Translation elongation factor EFG/EF2, domain IV (InterPro:IPR005517), Small GTP-binding protein (InterPro:IPR005225), Translation elongation factor EFG/EF2, C-terminal (InterPro:IPR000640), Translation elongation factor EFTu/EF1A, domain 2 (InterPro:IPR004161), Ribosomal protein S5 domain 2-type fold (InterPro:IPR020568), Ribosomal protein S5 domain 2-type fold, subgroup (InterPro:IPR014721), Elongation factor G/III/V (InterPro:IPR009022), Translation elongation/initiation factor/Ribosomal, beta-barrel (InterPro:IPR009000); BEST Arabidopsis thaliana protein match is: Ribosomal protein S5/Elongation factor G/III/V family protein (TAIR:AT5G25230.1); Has 61019 Blast hits to 57504 proteins in 4658 species: Archae - 1314; Bacteria - 38464; Metazoa - 2393; Fungi - 1714; Plants - 1333; Viruses - 1; Other Eukaryotes - 15800 (source: NCBI BLink).	MDDSLYDEFGNYIGPEIDSDQDDSDNDIPSDDDA DKRPSDDEAPSSAAAGPTNGWMTISHDEEDIEM VDNQVVLAEDKKYYPTAEEVYGDDVETLVEDED ELHLEQAIIKPVKALKFELGVKDSSTYVSTQFMLG LMSNPNLVRNVALVGQVHHGKTLFMDMLVEQTH HISTFDPKNERHMRYTDTRVDEQERRISIKSVPM SLVLEDSNSKSFLINIMDAPGHVNFSDEMSAALRL ADGAVLVVDAAEGVMVNTERAIRHATQERIPIVV VINKVDRLITELKLPPKDAYHKLRHTLEIINNHITAA SSTAGDVQLIDPAAGNVCFASATAGWSFTLLSFA KLYVKLHGVAFDANKFASRLWGDMYYHSDSRTF KRKPPATGGERSFVQFILEPLYKIYSQVIGEHRKS VEATLAELGVTLPNAAYRLNVRPLLRLACSTVFG TATGFTDMLVQHVPCGKEASSRKVDHIYTGPKD TVVYKSMENCDSSGPLMVNITKLYPKADCSVFDA FGRVYSGTIQTGQAVRVLGEGYSPEDEEDMTVK EVTKLWVYQARYRIPISKAPPGSWVLIEGVDASI MKTATLCNVDFDEDAYIFRPLQFNTLSVVKTATE PLNPSELPKMVEGLRKISKSYPLAITKVEESGEHT ILGTGELYLDSIMKDLRELYSEVEVKVADPVVSFC ETVVETSSMKCFAETPNKKNKITMIAEPLERGLA EDIENGVVSIDWPRKKLGEFFQTKYEWDLLAARS VWAFGPDKQGPNILLDDTLPSEVDKNLLNAVKDS IVQGFQWGAREGPLCDEPIRNVKFKIVDARIAPE PLHRGTGQIIPTARRVAYSAFLMATPRLMEPVYY VEIQTPIDCVSAIYTVLSRRRGHVTADVPQPGTPA YLVKAFLPVIESFGFETDLRYHTQGQAFCSSVFD HWAIVPGDPLDKGIVLRPLEPAPIQHLAREFMVK TRRRKGMSEDVSIQKFFDDAMVVELAQQAADLH LQMM
748	sp_078240_fxen	#N/A	-0.8858807	#N/A	uncharacterized protein	MKQGSRAPMYVGLAMTLMFISIMMPVVCGEETT LLGNLEGKNGNLRLKALFRADP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
749	sp_078890_ahzz	-0.893 14435	#N/A	#N/A	prolyl oligopeptidase family protein; FUNCTIONS IN: serine-type peptidase activity; INVOLVED IN: proteolysis; LOCATED IN: membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase S9, prolyl oligopeptidase, catalytic domain (InterPro:IPR001375), Peptidase S9B, dipeptidylpeptidase IV N-terminal (InterPro:IPR002469); BEST Arabidopsis thaliana protein match is: acylaminoacyl-peptidase-related (TAIR:AT4G14570.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MHPVGTKAGRKNTLKRSRSLSLDMPATVTTDNS MVSQKFAVDDDDSGEIPLFSVEEIVQYPLPGYVA PTSLSFSPDDNLITYLFSPDHSLSRKVFSFNLETG KHELVFSPVDGGLDESNISEEEKLRREFARERGIL GVTRYEWVKMLSSKKKIIMVPLPDGVYLYDLSSS KVDLKLQKTPSSPIIDPHLSPDGNMLAFIRGYELH VFDILYNKEKQLTFGAKGSILTHGLAEYIAQEEMD RKNGYWWSADSKFIAFTQVDSSEIPLYRIMHQG KSSVGPDAQEDHAYPFAGASNVKVRLGVVSIAG GPITWMDLHCGGNDDEEYLARINWMPGNILIAQV LNRAQTRLKVLKFDIRTGQRKVLMVEEHDKWVN LHDCFTPLDKSVSGGSGGFLWASEKTGFRHLYL HDVDGHCLGPITEGDWMVEQIAGVNEATGVLYF TGTVDGPLESNLYSVKLVPNRSGPLQAPVRLTH GNGKHIVVLDHQMQRFIDIHDSLTSPPKVVLCSLR DGSLITTLYDQPSGVPRFKRLKLEPPEIVQIKADD GTVLYASMYKPDETRFGPPPYKTMISVYGGPSV QLVSDSWINAVDMRAQYLRSKGVLVWKLDNRG SARRGLDFEAAVKHRFGRLDAEDQLAGAQWLIN KGLAKQGSIGLYGWSYGGYLSAMTLARFPEVFT CAVAGAPVTSWDGYDSFYTEKYMGLPSENLAGY EFGSVMHHVRNVKGKLLLVHGMIDENVHFRHTA RLVNAFVSAGKPYELLIFPDERHMPRRLRDRIYM EERIWEFIQRNL	
750	sp_079190_rnwe	-0.872 7959	-0.94 65337	-0.93 8546 36	histone deacetylase 9 (HDA9); FUNCTIONS IN: histone deacetylase activity; INVOLVED IN: histone deacetylation; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histone deacetylase (InterPro:IPR003084), Histone deacetylase superfamily (InterPro:IPR000286); BEST Arabidopsis thaliana protein match is: histone deacetylase 1 (TAIR:AT4G38130.1); Has 8723 Blast hits to 8511 proteins in 1453 species: Archae - 219; Bacteria - 3191; Metazoa - 1512; Fungi - 548; Plants - 472; Viruses - 0; Other Eukaryotes - 2781 (source: NCBI BLink)	MRSKDKISYFYDGDYGSVYFGPNHPMKPHRLCM THILVLSYDLHKKMEIYRPHKAYPVELAQFHSAD YVDFLHRITPDKQHHFMSEMAKYNLGEDCPVFD NLFEFCQIYAGGTIDAARRLNNKLCDIAINWAGGL HHAKKCEASGFCYINDLVLGILELLKYHARVLYIDI DVHHGDGVEEAFYFTDRVMTVSFHKYGDLFFPG TGDVKENGEREGKNYAINVPLKNGIDDASYTRLF KAIISKVVEFYQPGVIVLQCGADSLSGDRLGCFNL SIDGHAECVRFVKKFNLPLLVTGGGGYTKENVAR CWTVETGVLLDTDLPNEIPENEYIKYFGPDYSLRI PNGLIENMNSKSYLTTIKMQVLENLRCIQHAPSV QMQEVPPDFYIPDFDGDENPDERINQHTQDKLIQ RDDEYYEGDHDNDHHMDDV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
751	sp_079390_kohp	-0.852814	#N/A	-0.852819	cell division cycle 5 (CDC5); FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: defense response signaling pathway, resistance gene-independent, defense response signaling pathway, resistance gene-dependent, defense response to bacterium, defense response to fungus, regulation of transcription; LOCATED IN: chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF3351 (InterPro:IPR021786), SANT, DNA-binding (InterPro:IPR001005), Myb, DNA-binding (InterPro:IPR014778), Homeodomain-like (InterPro:IPR009057), Myb transcription factor (InterPro:IPR015495), Homeodomain-related (InterPro:IPR012287), HTH transcriptional regulator, Myb-type, DNA-binding (InterPro:IPR017930); BEST Arabidopsis thaliana protein match is: myb domain protein 3r-4 (TAIR:AT5G11510.1); Has 13137 Blast hits to 10813 proteins in 768 species: Archae - 14; Bacteria - 356; Metazoa - 3570; Fungi - 914; Plants - 4936; Viruses - 8; Other Eukaryotes - 3339 (source: NCBI BLink).	MRIMIKGGVWKNTEDEILKAAVMKYGKNQWARI SSLLVRKSAKQCKARWYEWLDPSIKKTEWTREE DEKLLHLAKLMPTQWRTIAPIVGRTPSQCLERYE KLLDAACTKDDNYEPGDDPRKLRPGEIDPNPESK PARPDPVDMDEDEKEMLSEARARLANTRGKKAK RKAREKQLEEARRLASLQKRRELKAAGIDPRQR KRKRKGIDYNAEIPFEKRAPAGFYDVAEEDRPVE QPKFPTTIDELEGERRVDREARLRKQDIAKNKIAE RQDAPGAIMQANKLNDPETVRKRSKLMLPPPQIS DQELEEIARMGYASDLLVGNQELAEGSGATHALL ANYSQTPRQGMTPMRTPQRTPAGKGDAIMMEA ENLARLRESQTPLLGGDNPELHPSDFSGVTPKK RDVQTPNPMLTPSATPGGAGLTPRIGMTPSRDG SSFGMTPKGTPIRDELRINEELDMHDSAKLELRR QAELKRNLRTGLSGLPEPKNEYEVVVRPAPEDT EEPEEKIEEDMSDRIARERAAEEARQQALLRKRS KVLQRELPRPPVASVELIRHSLMRAGEDKSSFVP PTDIEQADEIVRKELLSLLEHDNAKYPLNDKVQKK GSKRSANGVPIIEEFEEDVLKEAASFIKEEAQYLC VAMGHQEETLDNFVEAHNTCLNDLIMYFPTRNAY GLSSVAGTVEKLAALRNEFENVKRKMDDDNKKA QRLEKKIDVLTHGYKMRADKLKTQIELTFKQMNT GGTELDCFLALQRQENLAASYRIKGLWEEVQKQ KEVEKTLQKRYGERVAEQERVQQVMDQYRLQA QQLEEIEVKKRTEELAAFVAKEASFLKSIAPEPSS TTEVDDSSNFAVPKHVEEAPEVSLQNDIHVDGDV ASNADIKSNSLASETPLEDSSKSESEAAKEQNVDD NTNVVDGNLENSDEAAKEQNVDDNSNVVDGNLE NSDEAAKEQNVDNTNVVDGNLENSDEAAKEQNA GEAVENNGFVNGEPVIEAQGVDGKLGSQINKAE EGNETMTEDAGNFVESVENGASDNLV
752	sp_079840_xqus	-0.976628	-0.973341	-0.976883	Cytokine-induced anti-apoptosis inhibitor 1, Fe-S biogenesis; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF689 (InterPro:IPR007785); BEST Arabidopsis thaliana protein match is: Cytokine-induced anti-apoptosis inhibitor 1, Fe-S biogenesis (TAIR:AT5G18362.1); Has 539 Blast hits to 538 proteins in 227 species: Archae - 0; Bacteria - 0; Metazoa - 183; Fungi - 185; Plants - 75; Viruses - 0; Other Eukaryotes - 96 (source: NCBI BLink).	MAMLALTDDVILPLSIVSSSKNELNLEGPVLDNAL VITLASTLSALPLETSSMDTVIGVSKARDFPAVPL LAEMFRILKPDGRILIHLTELSGDLPFDNILSSLQR KILVAGFLEPEAGQKQPDAFSGVQSIQVKAKKPS WKIGSSFSIKASKGLPKVQINDDMDLIDEDTLLTE EDLKKPQLPPVGDCEVGSTRKACKNCSCGRAEE EQKVEKLGLTMEQLDNPQSACGSCGLGDAFRC STCPYKGLPAFKLGDKVSLSSNFLDADI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
753	sp_079920_jixy	-0.850701	-0.931246	-0.885115	26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit; FUNCTIONS IN: enzyme regulator activity, binding; INVOLVED IN: protein catabolic process, ubiquitin-dependent protein catabolic process; LOCATED IN: proteasome regulatory particle, base subcomplex, proteasome complex, nucleus, plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-like helical (InterPro:IPR011989), Proteasome/cyclosome, regulatory subunit (InterPro:IPR002015), Armadillo-type fold (InterPro:IPR016024), 26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit (InterPro:IPR016642); BEST Arabidopsis thaliana protein match is: 26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit (TAIR:AT2G32730.1); Has 1171 Blast hits to 1071 proteins in 297 species: Archae - 14; Bacteria - 33; Metazoa - 390; Fungi - 382; Plants - 143; Viruses - 0; Other Eukaryotes - 209 (source: NCBI BLink).	MFRRQSYGRPLLSILSPAQQDSSELVIASHTDPG VLTLLLQDQFGGLQVKHEGKWVNVKPIQVEYTV KLAASGFVFLKDLRPEEPEVLSLTDAPASASATT TNSDDADAPVQQGTVAAMAVDEEPQPPQPFEY TS
754	sp_081000_gwez	-0.939734	-0.987362	-0.98004	protein phosphatase X 2 (PPX2); CONTAINS InterPro DOMAIN/s: Metallophosphoesterase (InterPro:IPR004843), Serine/threonine-specific protein phosphatase/bis(5-nucleosyl)-tetraphosphatase (InterPro:IPR006186); BEST Arabidopsis thaliana protein match is: protein phosphatase X 1 (TAIR:AT4G26720.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSDLDKQIEQLKRCEPLKESEVKALCLKAIEILVE ESNVQRVDAPVTICGDIHGQFYDMKELFKVGGD CPKTNYLFLGDFVDRGFYSVETFLLLLALKVRYP DRITLIRGNHESRQITQVYGFYDECLRKYGSVNV WRYCTDIFDYLSLSALIENKIFGVHGGLSPSITTLD QIRTIDRKQEVPHDGAMCDLLWSDPEDAVDGW GMSPRGAGYLFGGSVVTSFNHTNNIDYICRAHQL VMEGYKWMFNHQIVTVWSAPNYCYRCGNVAAIL ELDENLNKEFRVFQAAPQDARGTPAKKPAPDYF L

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
755	sp_081150_szix	#N/A	-0.850332	#N/A	decapping 5 (DCP5); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: negative regulation of translation, deadenylation-independent decapping of nuclear-transcribed mRNA, cytoplasmic mRNA processing body assembly; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: DFDF motif (InterPro:IPR019050), FFD/TFG box motif (InterPro:IPR019053); BEST Arabidopsis thaliana protein match is: decapping 5-like (TAIR:AT5G45330.1); Has 14874 Blast hits to 10327 proteins in 603 species: Archae - 12; Bacteria - 902; Metazoa - 6418; Fungi - 2978; Plants - 1226; Viruses - 261; Other Eukaryotes - 3077 (source: NCBI BLink).	MAAEASRSATGSADSYLGSLISLTSKSEIRYEGVL FNINTDEASIGLRNVRSFGTEGRRKDGQQIPAND KVYEYIVFRGSDIKDLQVKSSPPVQTAPPVPNDP AIIQSHYAAPTASSTVPTASVPAAAGLLSDFNNQ MSQPGVPITTYQNSVPMYQPGGNLSPWNPSTP PTSNVSGVSVPTYWQGYYGPSGSQQMQQQSLL RPVQGLSMPPSVQPPSAQQPLQYPTMNSSLPT GASNISASQSMEFPPPLLQPIVSSTLSMQYNVHP DQSPVLSAESSTNRMPSMVPAQVAPVVPQSTKS PLLPPVLSNVLDKSTVGPLIPNQAKPISTTSMPFN SFSESLSAAGRKSTSPALVTPGQFLQPGPTVSQ TAQTEQKDVEVVQVSPLVSPPQPSTTSATTAAIV PGSAEAQAPLLPLPSPSNSKINEFPSQTRYVNVN RGGRGGRRGGNEFSRPVANFTEDFDFEAMNEK FKKDEVWGHLGKNKAKADDGLEYQDEYDDVS EIDSKPVYKKDDFFDTLSCNALDRESRNGRTRFS EQMRLDTETFGNFPINRGFRGGRGPGGGRSRG GYYGRGYGYGYVGRGRGHNNVYRAI
756	sp_081630_zsfs	#N/A	#N/A	-0.852177	ETHYLENE-DEPENDENT GRAVITROPISM-DEFICIENT AND YELLOW-GREEN-LIKE 3 (EGY3); INVOLVED IN: response to high light intensity, response to hydrogen peroxide, response to heat; LOCATED IN: chloroplast; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 9 growth stages; BEST Arabidopsis thaliana protein match is: Peptidase M50 family protein (TAIR:AT5G35220.1); Has 660 Blast hits to 652 proteins in 203 species: Archae - 107; Bacteria - 240; Metazoa - 53; Fungi - 13; Plants - 117; Viruses - 1; Other Eukaryotes - 129 (source: NCBI BLink).	MATLFISPPSSSWVSNKTKKLASFSSNSHFHHPV TLSSYSKPSKPRKLRFSQQDEGERESSSVVVVD EKSKNDEDTQNGVEEREEVDDEQKQQELDWKN DAEFKKFMGNPSIEAAIKLEKKRADRKLKELDDE SKSDNPIVGVFNRIAPENLLREKETLEKVEQAFKA LDLNKVCLVQPKSEIDLQFESTNLSTPWGYLSAIA LGVTTFGTIALTSGFFVKPGATFDDYLANVVPLFG GFLTILGVSEIVTRLTANRYGVKLSPSFLVPSNWT GCLGVMNNYESLLPSKKALFDIPVARTASAYLTS LALAVVAFASDGSFNGGDNALFVRPQFFENNPLF SFIQYVIGPYADDLGNVLPYAVEGVGVPVDPLAF AGLLGMVVTSLNLLPCGKLEGGRIAQAMFGRNT ATLLSFATSLLLGIGGLSGSVICLSWGLFATFFRG GEENPAKDEITPLGDDRYAWGVVLFLICFLTLFP NGGGTFSSSFFTGPFFRGDL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
757	sp_082110_zxgk	-0.888367	-0.952638	-0.911709	Phox-associated domain;Phox-like;Sorting nexin, C-terminal; CONTAINS InterPro DOMAIN/s: PX-associated, sorting nexin 13 (InterPro:IPR013996), Sorting nexin, C-terminal (InterPro:IPR013937), Phox-like (InterPro:IPR001683), Phox-associated domain (InterPro:IPR003114); BEST Arabidopsis thaliana protein match is: Phox-associated domain;Phox-like;Sorting nexin, C-terminal (TAIR:AT2G15900.1); Has 944 Blast hits to 779 proteins in 177 species: Archae - 0; Bacteria - 4; Metazoa - 546; Fungi - 164; Plants - 152; Viruses - 0; Other Eukaryotes - 78 (source: NCBI BLink).	MSSQRQLVTARDLVEEAKKRIVILSICVVGLSYLM SLTSTSVWWNLPAAAVLIIIFRYFSLDFETRRKEA AYNSRNSGKITIQEKPLDVPKTLVKKDQWRSKVN SPPVEDAIDQLTRHIVSEWVTDLWYSXGEVLKEL VQLMNGVLEEVAYCMKHINLIDLLKRDVIHLICKHL ELFRVCQAKIKKNLASLSFDQRDMELKIVLGSEN KLHPLLFSIESEHKVLQHIMDGLLSLTFKPEDLQC SFFRYVSRELLACAVIRPVLNLANPRFINERIEQV VISARKKNGEAATTNNSSLLKQNGGNSDHFSPFP DPSVSGVELVELKSIQSKSSMDKLGKEKVDSEDA SKDPLLLVDTRSTRSWSSLPFGSAAAAGSHIDRS ASGGEWGDMLDYFSRKKKQTLAPENFENMWSK GRDYNGKDTQITNHVHPGSSIKRSDAVNSSKVFA KQKERHNIASAERGVILPGSNVRPQSETMYISAD SNMSTPPPLTSDEDDEEQKLMETELLSSSQSSS GDEEDRVTGLDSPTTKVWDGRTNRSSGVSHIRH PLEGLEGKKGKKGNRGQYQKLHRQQSGRKRSR LGSQKVPLWQEVERTNFLPGDGKDVHPLKPNVK DDDSSDDYDMEIVGRLQSGTTASSSASSLPGSF TSTVNSPQTAFLDDTFFKLRCEVLGANIVTSSSKT FAVYSISVIDVERNSWSIKRRYRHFEELHRRLKEY PEYNLHLPPKHFLSTGLDVLVIQERCQLLGQYLK KLMQLPTVSGSIDVWDFLSIDSQTYTFSSAFSIIET FSVKPHDSSANMSSLGGPKVNHLQPRRDHAESE NKDQSPPKNPVKEVGKLAGDSDSRIQSNGRKGN EGAVQQNVGKEGDIADDPTFPMEWVPPNLSGPI LDLVDVVFQLQEGAWIRRKAFWVAKQVLQLGM GDALDDWLVEKIQRLRKGSVVASGIKRVEQILWP DGIFITKHPSRQRPPASGSPSQKSSDGKIPDQLS SPKAENFEDIAELQEEEADRRAKFVYELMIDNAP AAVVSIVGHKEYEQCAKDLYYFLQSSVCLKQLVM DLLQLVLLHAFPELDHVFKQLDEEKHKFGEFKAT
758	sp_082510_cuwe	#N/A	#N/A	-0.86476	uncharacterized protein	MKILAVLLPPSQHVLAFRLPAIAAAPAKSVQYRGA SVAPLKNYMADVALKLHSSSACYLNTEYS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
759	sp_083500_skgc	-0.950976	-0.87783	-0.862477	DWARF 4 (DWF4); CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, conserved site (InterPro:IPR017972), Cytochrome P450, E-class, group I (InterPro:IPR002401); BEST Arabidopsis thaliana protein match is: Cytochrome P450 superfamily protein (TAIR:AT5G05690.1); Has 31120 Blast hits to 31036 proteins in 1650 species: Archae - 61; Bacteria - 5303; Metazoa - 10591; Fungi - 5722; Plants - 7996; Viruses - 3; Other Eukaryotes - 1444 (source: NCBI BLink).	MSDLEFFLYLLPFVLALSWTLNFLKRKKPQPNLP PGNSGLPFIGETFGYLKPHPATTIGKFMEQHIERF GKIYRANLFGEPTIVSADAGLNRFIFQNEGRLFES SYPSSIGGILGKWSMLVLNGDIHRDMRNISLNFLS QSRLKAHLMKEVERQAVFVLNTWQEDCSFSAQ DEAKKFTFNVMAKQILSLDHGEAETEHLKKEYINF MKGVVSAPINLPGTPYRRALKSRTAILDFIEMKME EKLKKKKMNDGSENIDEDDLLGWVLKHSTLSKE QILDLILSLLFAGHETSSVSIALAIFFLEGSPSAVEQ LREEHLLIARAKQQSAEKELNWEDYKKMEFTQC VISETMRLGNVVRFLHRKALKDVRYKGYDIPCGW KVLPVLAAIHLDPAFYDQPHVFNPWRWQQNGIN RGGVSPSATASNSNSNSSPTSGNNFMPFGGGQ RL
760	sp_085060_arwu	#N/A	#N/A	-0.864728	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; Has 428 Blast hits to 428 proteins in 59 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 428; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MASSSSTAQNPRHLLNHLYGGPTSAPPPTPSTQ PSLSSSTSDAAADTLSRLLHRLPPSLSLPTRLSPR APAKSSPHLPTLSLSSTSFPSSGVDSNLFELGYF QLTHHSITPKLAESAESESLSLFDLPLDEKLSSFP TDWPLGFNNDDDDDDDNDGDDHQTGSFFLDGQ CSATNTAELSLTSLRELTRELEKLGLEVVEKLASA VGFENPFREGGVVGSNKQSSLMWVTSSDGDKA RRVYPYVVGLQYQIRPRKWDLVSDSGSVSVEPQ VGSVLVTLGDIAQVWSNGKLKKVRGRPAPAFGDI KSNISMTMLLTLPLESTVSPLLLVPNYTGEKEANN QQKDKDNQTINPNTNSNDGNDVDNEDQRIFNSF SFEDYAWRVYHERLLFKDPLDRYRVIH
761	sp_085280_uxwe	-0.881942	#N/A	#N/A	DA1-related protein 2 (DAR2); FUNCTIONS IN: zinc ion binding; LOCATED IN: chloroplast; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, LIM-type (InterPro:IPR001781), Protein of unknown function DUF3633 (InterPro:IPR022087); BEST Arabidopsis thaliana protein match is: LIM domain-containing protein (TAIR:AT4G36860.1); Has 1539 Blast hits to 1276 proteins in 109 species: Archae - 0; Bacteria - 10; Metazoa - 1178; Fungi - 26; Plants - 170; Viruses - 0; Other Eukaryotes - 155 (source: NCBI BLink).	MADATQIHHLSQPCIYGLSAPPNTEKKSRFMKWL SKIFKGSGGSGGGSSRSSSAGGRHPQFLEDEN MVWRAPSRSLDDRPRASKDQEELDRALAFSLSD DLKRPGSCKRRTDNDENTAKGLQDSLHRPSYTP YMPSYPDYPPYDPVPFYPRGYRLCSGCNGEIGY GNYLGCMGTFFHPECFCCRSCSYPITEHEFSLS GKDPYHKTCFKELTHPKCEVCHQFIPTNTAGLIE YRCHPFWSQKYCPSHEHDQTARCCSCERLESW NTRYIALGDGRSLCLECMESAIMDTGDCQPLYHA IRDYYEGMNMRIDQQIPMLLVERQALNEAIVGEK NGFHHMPETRGLCLSEEQTVTSILRRPRIGGHRL VGMRTQPQKLTRRCEVTAILVLYGLPRLLTGAILA HELMHGWLRLKGYRNLSPEVEEGICQVLSYMWL ESEVMPGFKNMASTSTASTSSFSTSKKGKTSDV EKKLGEFFLHQIAHDASPAYGGGFRSANAAVNKY GLRRTLDHIRLTGSFPI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
762	sp_085410_xpfa	-0.912433	-0.949762	-0.889079	DEFECTIVE KERNEL 1 (DEK1); FUNCTIONS IN: cysteine-type endopeptidase activity, calcium-dependent cysteine-type endopeptidase activity; INVOLVED IN: cell fate specification, embryo development ending in seed dormancy; LOCATED IN: plasma membrane; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase C2, calpain, domain III (InterPro:IPR022683), Peptidase C2, calpain, large subunit, domain III (InterPro:IPR022682), Peptidase, cysteine peptidase active site (InterPro:IPR000169), Concanavalin A-like lectin/glucanase (InterPro:IPR008985), Peptidase C2, calpain, catalytic domain (InterPro:IPR001300), Peptidase C2, calpain family (InterPro:IPR022684); Has 3377 Blast hits to 2425 proteins in 256 species: Archae - 0; Bacteria - 87; Metazoa - 1580; Fungi - 264; Plants - 115; Viruses - 0; Other Eukaryotes - 1331 (source: NCBI BLink).	MESEQHHIVLACVVSGALFALLGSSSFSILWVVN WRPWRLYSWIFARKWPDLFQGPRLGVICGALTL LAWFVVVSPLLLLVGWGSWLILILGRDIIGLAVIMP GTALLLSFYSIMLWWRTRWQSSRAVAFLLLFAV ALLCAYELCAVYVTAGSSASERYSPSGFFFGVSA IALAINMLFICRMLFNGNGLDVDEYVRRAYKYAYS DCVEVGPVTCLPEPPDPNELYPRQSSRVSHLGT LYLGSLLVLLVYSILYGLTAKEARWMGAITSAAVIV LDWNMGACLFGFQLLKSPIAALFIAGASRIFLICF GVHYWYLGHCISYLVMASVLLGAAVCRHLLVTNP SAERRDALQSTVIRLREGFRRKEHNTSSSSSEG CGSSMKRSSSAEAGHLGHGVESVGRNGGHCTG EAVGWNNGTLLRTTSSHEGINSEKSMDSGRPSL VLRSSSCRSVVQDPDVGPSFVEKSDHHSVVGYS SGGLESQGYDSSTSTSVNYLPDLNLALVYQEKLN DPRIAAILKKRAKQGDRELTNLLQNKGLDPNFAM MLKEKGLDPTLLALLQRSSLDADRDHCDNTDITIV DSNSVGDALPNQISLSEELRLRGFEKWLQILRSL WHFVAGTPERALACFSSVFILETVGVAFSRPRVI KLVNATHQQFEFGICVVLLSPVICLIMAYLRSLQA EDTSTSSKPRKYGFFAWLLSTCVGLLLSILSKSS VLLGLSLTVPLMVACLSIAVPLWIRNGYNFWIVQD NPAPRTENHQALGRIEGIAFLVCVAVFVGSVISLG VIVSFKPLDELSYNGWNADGKIYASPYASPIYLG WAMASGIALLVTGVLPVISWFATYRFSVFSGVCL VLFSVVLVAFCGVSYAEITNSRIDHIPTNGDFLAAL LPLLCIPAILSLCTGLHKWKDDEWRLSCSVYVFITI GLLLLLCAISAVIVIVQPWTVGVAFLLVLVLIVLAIG AIHLWASNNFYLTRAQMFFVCFLAFLLAVAAFLV GRFEAKPFVGASVGYFSFLFLLAGRALTVLLSPPI VVYSPRVLPVYVYDAHADCGKNVSGAFLMLYGIA LATEGWGVVASLNIYPPFVGAAVSAVTLVVSFGF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
763	sp_086420_utqj	-0.92359	-0.929062	-0.959454	Protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase family protein (TAIR:AT5G18190.1); Has 16177 Blast hits to 16113 proteins in 1419 species: Archae - 20; Bacteria - 5310; Metazoa - 4811; Fungi - 1519; Plants - 1815; Viruses - 223; Other Eukaryotes - 2479 (source: NCBI BLink).	MPELRSGARRSKRLGDLQQPAPELNDQGENFLP STQNKTRRRGGGRGRGANAAAGVAKAAIGARQ TGGAGRGRGVRVIDLDPEIPGDGLPQVAPIGVAG ATHNRVEGAAGKDIAMEGGSGDKLVAAEDDSNL TPVPERVQVGTSPVYKIERKLGKGGFGQVYVGR RMVGGTDRIGPDAIEVAVKFEHRNSKGCNYGPP YEWQVYNSLNACYGVPSVHFKGRQAEYYILIMD MLGPSLWDVWNSLGQSMSPNMVACIAVEAISIL EKLHLKGFVHGDVKPENFLLGQPGTADEKKLYLI DLGLASRWKDVSSGHHVEYDQRPDVFRGTIRYA SVHAHLGRTGSRRDDLESLAYTLIFLIKGRLPWQ GYQGDNKSFLVCKKKMGTSPELMCCFCPAPFKQ FLEAVTIMKFDEEPNYAKLISYFDSLIEPCAPLRPI KIDGALKVGNKRGRMLVNNLEEDEQPKKKIRLGS PATQWISVYNARRPMKQRYHYNVADARLKQHID KGNEDGLYISCVASATNLWALIMDAGTGFGAQV WELSNVFLHKDWIMEQWEKNFYISSIAGASNGS SLVVMSKGTSYTQQSYKVSDSFPFKWINKKWKE GFHVTSMTTAGSRWGVVMSRNAGFTDQVVELD FLYPSEGIHRRWESGYRITSMAATPDQAAFVLSI PKRKTMDETQETLRTSAFPSTHVKEKWAKNLYIA SICYGRTVC
764	sp_086510_tjgw	#N/A	#N/A	-0.890728	cytokinin oxidase/dehydrogenase 1 (CKX1); FUNCTIONS IN: cytokinin dehydrogenase activity; INVOLVED IN: N-terminal protein myristoylation, cytokinin catabolic process, meristem development; LOCATED IN: vacuole; EXPRESSED IN: lateral root, shoot apex, hypocotyl, root, flower; CONTAINS InterPro DOMAIN/s: Cytokinin dehydrogenase 1, FAD/cytokinin binding domain (InterPro:IPR015345), FAD-binding, type 2 (InterPro:IPR016166), Oxygen oxidoreductase covalent FAD-binding site (InterPro:IPR006093), FAD-linked oxidase-like, C-terminal (InterPro:IPR016164), FAD linked oxidase, N-terminal (InterPro:IPR006094); BEST Arabidopsis thaliana protein match is: cytokinin oxidase/dehydrogenase 6 (TAIR:AT3G63440.1); Has 6769 Blast hits to 6763 proteins in 1376 species: Archae - 168; Bacteria - 3882; Metazoa - 142; Fungi - 1302; Plants - 645; Viruses - 0; Other Eukaryotes - 630 (source: NCBI BLink).	MLLKVFLVLLLNLAEHNINHCSYLPHATPDFTNLS SLELVSSLKKLNLNGNFDFENTRRVANDFGNRYH FLPMAVLLPKTVSDISTTIKHVINMGHNSELTVTA RGHGHSLQGQSQSHGGIVISMESLEEPKMHVHT GEQPYVDVSGGELWINILHETLKQGLSPKSWTD YLYLTVGGTLSNAGISGQAFRYGPQINNVHRLEIV TGKGDILNCSAKENADLFYGVLGGLGQFGIITRA RISLEPAPRSETEHLLSKLNYISHTLFQSEVSYVN FLNRVHLSEIKLREKGLWDVPHPWLNLLVPKSKI YSFAEEVFGHILTDTSNGPILIYPVNQSRWNHKTS MVTPSEDVFYLVAFLTSAMPSSSGSDGLENILTK NKRILDFCEMAKLGAKQYLPHFDTQDGWRAHFG PRWDVFARRKLTYDPLSILTPGQRIFRKGRAIS
765	sp_087460_pewf	-0.901059	-0.932983	-0.954461	uncharacterized protein	MKAMMEDDQIDGDGDEIDEGDDAGIEDEEDLIVM LRRRLLGQKWKKARVELLSLWKWLNPGVRVYR KMRK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
766	sp_087470_gmaz	-0.8808831	-0.9237714	-0.9831888	Noc2p family; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised protein family UPF0120 (InterPro:IPR005343); BEST Arabidopsis thaliana protein match is: Noc2p family (TAIR:AT3G55510.1); Has 4709 Blast hits to 3397 proteins in 331 species: Archae - 4; Bacteria - 179; Metazoa - 1693; Fungi - 681; Plants - 249; Viruses - 70; Other Eukaryotes - 1833 (source: NCBI BLink).	MRCFRMLNCIAAATDTFIPVSMLLLDMLEMKELN CPPSGGVSKAVDLRTILKEACVSSVIDELGEHLA QRSYSVAFFELSFIPAVRLRNFCKCTKVERCRKD MKELNRQIEANSEFTNKKRATISFLPSEPAISSFL EEEKQSGGYPLSKYVATLRQKAQQKNVSVSESS VLYGAEPPISDKMLAESDDEESDDEEDGDKGEA VFNSFWTQQKETRDKPPPEDEKTQKSKKKSRSS QKDAIDEDIVEYLVLSSYEYDEPMDDSPPLEETVK TEIGISKRRNKRKPTIEEKWFNEE
767	sp_087600_wiun	-0.8808862	-0.950002	-0.9823493	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT1G68980.1); Has 11641 Blast hits to 5218 proteins in 154 species: Archae - 0; Bacteria - 0; Metazoa - 18; Fungi - 47; Plants - 11427; Viruses - 0; Other Eukaryotes - 149 (source: NCBI BLink).	MWRRRAFSLLRPPQTRPFSTTTTTTATGETPSL YSFLQPSLFSLKPQQTASPPSPPPTSSKIITPEHK TTLEITLQQSLSTADTDEAWKSFKSLTNTSSYPS KSLTNSLITHLSSLSDVHNLKRAFTAVVFLVEKNP NLLDFDSVSRVLSSMAAANYAAPAFSLVKCMFKN RFFVPFDMCGPVLVEISKKNGSFVGYLGVFHENC RVAIDEKLDFMKPNLDACNAALEGCCSDLESVEE AERVIELMSVLGVRPNELSFGLLGYLYALKGLEG KIVELGSLMGGFGFTDKMVFLSYLIRGYVKAGSM ESVSKTVLCSLKAGEGDGSKFSEETYCEIVKGFL KHGSIKDLAALVAEAQMSEPENVVGEQSTGFGIV NACTVLGMLDKAHSILDEMKAKGAVVGLGVYLSI LKAYCKEYRTAEAAQMVAEINCSGLHLDVSCYDS LIEASMSSQDFQSAFSLFRDMRDARISELKGSYL TIMTGLTENNRPELMAAFLDEVVEDPRIEVGTHD WNSIIHAFCKAGRLEDASRTFRRMAFLHFGPNEQ TYLSLINGYVVAGKYFNVLMLWNEVKRRVSTDDY GGFKLNHKLIDAFLYALVKGGFFDAVMQVVDKSR EWKIFIDKWRYKQAFMETHKKLKVSKLRRKNFKK MEALIAFKNWAGLSV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
768	sp_088110_anou	-0.910845	-0.884044	-0.873172	Helicase/SANT-associated, DNA binding protein; FUNCTIONS IN: DNA binding; LOCATED IN: apoplast; EXPRESSED IN: male gametophyte, cultured cell, pollen tube; EXPRESSED DURING: M germinated pollen stage; CONTAINS InterPro DOMAIN/s: SANT, DNA-binding (InterPro:IPR001005), HSA (InterPro:IPR006562), Homeodomain-like (InterPro:IPR009057), HAS subgroup (InterPro:IPR013999), Helicase/SANT-associated, DNA binding (InterPro:IPR014012), MYB-like (InterPro:IPR017877); BEST Arabidopsis thaliana protein match is: Helicase/SANT-associated, DNA binding protein (TAIR:AT3G24880.1); Has 9846 Blast hits to 7198 proteins in 452 species: Archae - 0; Bacteria - 514; Metazoa - 4085; Fungi - 1476; Plants - 1172; Viruses - 23; Other Eukaryotes - 2576 (source: NCBI BLink).	MRGFSTGSTFLVNAEVDSMGGVVDGGVGTGMK PSPRRAAIEKAQADLRVEYDVREERRRELEFLEK GGNPLDFKFGHAASVSVQSTSHTYQHPDQILTS EAKGSFALTASPHGDSVESSGRLGVTTLCEPNS ADNFDGENEILESERKSKHPTRGTITPSEHSSQL DGSQNVKESEDSPIFHPKKGQAYRRRNRSRTNR DGPRSSSSDMASRGQNSSLPGRHAILAKPVASE YQFELESEGTHGLHTSTCSPASEISEGKLTISTSK ENREELQNSSVKADEIPAAATNVEPDIVAGKETV VPVDAGGPPCSGTEKKEDIVSASRLDESGKFNG DGKPTPNDAIISNDAVVTKGLDSESSCTRTSRSV DGNGNIEIDLYTNLKNVDSNGVPKEHKLPLGEKCI NVDDEMLSEKHVTNPLDSVSCIKEDQHSISHHGN GFADKDQEVDRDRYPSTDKPDEPINREELLALDH AETKPDKDHGHDSVQKVESSCPVRPLSSVEPSS HEVNRNDLSAPPPASNTQTSRENQRKQLDKAHE DRVLEEARIIEAKRKRIAELSVGLTPREHRRKSQ WDFVLEEMAWLANDFSQERVWKLTAAAQICHR AAFASRLRLEKQNRKSWKQKSVSLTLAKAVMEFW HSAGLRVENVTHNASSGADKTGQVKDVAGGFE KDDVSEKKLQSTGCKRDLAVYGYAARFLQYNRS SPPFVKVEGSGTPDVACDIDLIGVSWQDQLTEEN LFYLVPPGAIGIYRKSIEQCLSDFERVSNNIQEEV DTSLYDAATEIGYQENDYEEEDEGENGYDLPFDS GRSSKKKRKNMKSYEYGGDFGYVRSLDSRNGT PQSALVGKRPANNHGVGSIPTKRIRTASRPRVVG PFGAGSAGVTLAPSRADASSGDTNSFQDDIVPR GSEVDSGMGFENQSSFDSPEISAKPKKKKKIKHP GAGYDPNWQVDSSLHNEQKDHARKRYESYQLE PNGNSGMYAQHTKKPKTFKHPLDNPFDSTLPMA GSVASPVASQMSNMSSQSKIIKLIGGRDRSRKAK GLKVPAAQTGSGSPWTSFEDQALIVLVHDMGPN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
769	sp_088380_sjcj	-0.8561148	-0.8044455	#N/A	Pentatricopeptide repeat (PPR-like) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT3G06920.1); Has 55395 Blast hits to 14888 proteins in 319 species: Archae - 4; Bacteria - 71; Metazoa - 778; Fungi - 987; Plants - 51638; Viruses - 0; Other Eukaryotes - 1917 (source: NCBI BLink).	MFLVPNPVSDSRISLPSRKSSARMLDERFIRILKIF KWGPDAEKALEVLKLKVDHRLVCEVLKIDVDINV KIQFFKWAGKRRNFEHDTTSYMSLIHCLLEAGSA GELWKTIQDFVRSSCHFSPSEFSEVLRVLGHAK MLSKAISIFYHIKRYKCKPTSNTYNSLILMLMQGG HYEKVHELYNEMCNDDSCFPDTVTYGALIQAFG KLGRDDSAVRLFDEMKENGLVPNAKIYTTVLAIYF KLAKADKALDLVKEMKDKGCSLTVYTYTELIKGL GKLGRVDEAYDTFLSMLKEGCKPDVVLINNLINVL GKSGLLSEALKVFEDMKSLQCIPTVVTYNTVIKCL FEANAPAAQASALFEKMKAIGIAPSSYTYSIMIDG FCKKNRVEKALMLLEEMDEHGFPPCPAAYCSLIN SLGKAKRYDAAKELFQELKETSGTSSARVYAVMI KQFGKSGCLDEAVDLFNEMKKLGCSPNVYAYNA LMSALARAGMINEAQMLLRTMEENGCAPDINSQ NILNGLARTGGPKQAMEMFEKMKDLSIKPDAVS YNTMLSCLSRAGMFEESVKLMKEMSSNGFKYDQ ITYSSLLEAVGKVDDDPIPPTFSQ
770	sp_088430_fcnq	#N/A	#N/A	-0.8553835	kinase associated protein phosphatase (KAPP); CONTAINS InterPro DOMAIN/s: Protein phosphatase 2C, manganese/magnesium aspartate binding site (InterPro:IPR000222), SMAD/FHA domain (InterPro:IPR008984), Protein phosphatase 2C-related (InterPro:IPR001932), Kinase associated protein phosphatase (InterPro:IPR016660), Protein phosphatase 2C, N-terminal (InterPro:IPR014045), Protein phosphatase 2C (InterPro:IPR015655), Forkhead-associated (FHA) domain (InterPro:IPR000253); BEST Arabidopsis thaliana protein match is: HOPW1-1-interacting 2 (TAIR:AT4G31750.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	VDEERPLVQEDSDQLLSQRNEPLKFHTSECSHL QTEGLFNLPGSHSLTYKQRLPPPAPHSTLGGSLV LDIGNQAEDLLVGQTLRRPLLTNHFIEEQKCTTKD GLNHDSRLRPDNVKFQSISKDVLDQGSCLSLEVI TGPACGLQSSVWSTDTSKLPLTLGRVPPSEVVL KDSEVSGKHAMIKWNSNKLKWELIDMGSLNGTL LNTKAVHHPDFGSRTWGEPKELVSGDVITLGTT SQISVQITPLAQSEVPVGVGIASDPMSTRRGGKR LPMEDVCYYQWPLPGAEQFGLFGICDGHGGAA AAESVSKILPETVATILLDSLRREKVFSRQDASDV LREAISQTEACLNHYYEGCTATVLLVWGDGYDSF FAQCANVGDSACIMNIDGRQIKMTEDHRITSYSE RQRINKTGNPLKDGDTRLCGLNLGRMLGDKFLK EQDARFSSEPYISQVVPIGQASEASVVLASDGLW DVISAKKANQFVLQAKERYSNDRKACAQKIANFL LNEARKLRTKDNTSVVYVDFDSALRSLCKYEP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
771	sp_088790_dipa	#N/A	#N/A	0.8782224	binding; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: CCAAT-binding factor (InterPro:IPR005612), Armadillo-type fold (InterPro:IPR016024), Nucleolar complex-associated (InterPro:IPR011501); Has 3184 Blast hits to 2630 proteins in 361 species: Archae - 21; Bacteria - 280; Metazoa - 1055; Fungi - 428; Plants - 179; Viruses - 26; Other Eukaryotes - 1195 (source: NCBI BLink).	MGKKQKQKQKVTLPPALPPDVPPDEDIVPSDEDV EFVDKNPAYASFVSRLDTQSITKHVNRVANAKED DLEALYEKFLRRKTLENEEKKGIEVDRVDALPIKT LDGQLYYKTVEKEAKKPETSDKEDESDEDKGAF KLTKAERRAKLKKKKKDEKKQAKEPVKTEEEPKD PQSEVLAEVREDLTAAEVAETKQHKIAELGTALL EDPEANIKSLKELLSFCNDDDKAIAKAALLSSLVV FRDIIPGYRIRLPTEKEQEMVVSKAVKKLRYFETT LLLSYKAYLQRLISLQPLPEYNRVAVRCLCTLLDA HPHFNFRENVIPAVVKSLASADDVVRKLCSTTIQS LFKNEGKHGGEATIESVQLIADLVKIRNCQLHPES IEVFLALSFDEDLARSKKAAEDDEKRKNKKYKKGK KSEEQVQTAVSDKKRSKKELLSKTRDEVNADYK AASLALDEKERRRLQSRTLSAVFNTYFRILKHTIQ PDATRFKANATASGASLGGHPLLSSCLKGLGMF SHLIDIDFMADLMKHLKILASGTNLGNSDSSLTVS ERIDCCIVAFKVMRKNLDALNVDLQDFFIQLYNLIL EYRTERDQGEILAEALKIMLCDDRQHDMQRAAAF IKRLASSSLCFGSAESMAALVTAKNLLQKNFKCR YMLENDVGGGSVSGSVAKYQPYASDPHLSGALS SVLWELNLLSKHYHPTVSAMASSISKINTSSNQV YMPTVSPQQAYNELLLEKEQFSIKSDIKLSNKRK RGNPNLLGEDESSLEVDEEALRKKLSEHYTVVH DIVENEKLRRELGLTTLSLQVHEEYKKVKKIKQEG SKQKKRKQ
772	sp_089300_mepf	#N/A	-0.8363263	#N/A	Protein of unknown function (DUF620); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF620 (InterPro:IPR006873); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF620) (TAIR:AT1G49840.1); Has 219 Blast hits to 218 proteins in 2 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 219; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MERKQGFFSALKEEVVRGLSPGKSRARSPARSG SAMTGLLRRRKGHRVAIPEVLIARSGSLRPPEML SPLREGPDQCGAREERDEGGWAHWMKRNPSIS GSSSSSSAAACKRSDLRLLLGVLGAPLAPVHVSA SEPLPHLSIKDTPIETSSAQYILQQYTAASGGLKL QNSIHNAYAMGKLKMLASDIETPNKVIKNRSSSK SAESGGFVLWQMNPDMWYVELALGGSKVHAG CNGKLVWRHTPWLGAHAAKGPVRPLRRALQGL DPRTTAAMFANARCIGEKKINGEDCFILKLCADPK TLKARSEGPAEIIRHVLFGYFSQKTGLLVHIEDSH LTRIQSNGGDAVYWETTINSCLDDYRPVEGIMIA HSGRSVVTLFRFGETAMSHTKTRMEEAWTIEEV AFNVPGLSIDCFIPPAEIRYSSVGDSCELPHGDRL KTMSATAAAYRAKVASMDRSYETNTSNMTWQM DNMHV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
773	sp_089350_jiky	-0.918975	-0.958023	-0.935324	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein; FUNCTIONS IN: DNA binding, zinc ion binding; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, PHD-type, conserved site (InterPro:IPR019786), Zinc finger, PHD-type (InterPro:IPR001965), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Acyl-CoA N-acyltransferase (InterPro:IPR016181), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein (TAIR:AT1G05380.2); Has 1402 Blast hits to 1245 proteins in 121 species: Archae - 0; Bacteria - 0; Metazoa - 811; Fungi - 61; Plants - 450; Viruses - 0; Other Eukaryotes - 80 (source: NCBI BLink).	MEGRNRSGSHSGGVVKNHGSSGCLIIKKKADGG SGVGSSKGRKLPESKEKKRSRLNASESVSSDEE LLEFYPRRGVSGMNQVSGGASFGYESNDFEDVS FGRNGLMREAGPKRSKLDVYDFDEYDDLSDAEM MRMRHYHDQGIGGEERRFIRAMPGMNNGSGLD FESGPSKGDIIYTRRNSFHGGSSSKKINGGDLYS GKRRFEMSSELPVSMGKEKLGGSDEAIRLQGKN GVLKVRVKHKEEISEQFSKKYVHVKGNNVALRS QEPASKNIVARPSPYSEKNVSKKPIAVDRLDKTQ RKKRKPPESSEESDEEANDSDSSPKLQTDCVQV HRSMKRGRSGGSTESPQSAQPIKLKEGKVKRGS GTEKQLLREKIRNMLLTAGWTIDYRPRRGRDYSD AVYVNPTGTAYWSIIKAYDALQKQLEEEGNSIKV GDDCPSFAPISENVLSKLTRQTRKKMEKELKRKK REEAGGKRKKEFTSGSFEEDESDTDSGGSGRQ DDKLSSFMKRSGKSLKGASKESCRDRSNSNSKM QERGEKSSSMSNSKFIHARKSKKIGRCTLLVRNS GKGLSSESDGFIPYTGKRNLLSWLIDSGTVNLSE KVHYMNRRCTRVMLEGWITREGIHCGCCSKILTV SKFEIHAGSKLRQPYQNIYLSSGASLMQCQVDA WNRQKDSLRNGYNKININGDDPNDDTCAICCDG GNLICCDSCPSTFHQSCLGIEKLPPGDWHCPHC SCKFCGLADMSVTGENNKRQDPVLTCSFCEGKY HCSCVEGMDGTLINLNSPPFCGKTCQELSEHFQ KLLGVKHEMEAGLFWSLIHRTDLDSDTSVRGYP QRVEWNSKLAVALNVMDECFMPIADRRSGINMI QSVLYNSGSNIARLDYSGFYTAILEKGDEIIAAACI RIRSTQFAEMPFIGTRHIYRRQGMCRRLFSAIEAA LRSLKVEKLVIPAISELMDTWTTIFGFSSLEELDK QDMKLGNVLVFPGTDMLQKKLLNQDSINRDLSV TGADSFETDQEASVMLGLRKKSDIDVPPGNRSV APEPENEEMKGKPVTEVCSQDTVCTDDNSLHSA
774	sp_089480_dnfr	#N/A	-0.88037	#N/A	uncharacterized protein	MLSFFRCRLASELYMLVCSSALQFKIGLRGALAS SASLKCLRPKLTC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
775	sp_090650_yxdc	#N/A	-0.976727	-0.693386	ARABIDOPSIS THALIANA CATION/H+ EXCHANGER 18 (ATCHX18); FUNCTIONS IN: monovalent cation:hydrogen antiporter activity, sodium:hydrogen antiporter activity; INVOLVED IN: cation transport; LOCATED IN: integral to membrane; EXPRESSED IN: 10 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Cation/H+ exchanger (InterPro:IPR006153); BEST Arabidopsis thaliana protein match is: cation/H+ exchanger 17 (TAIR:AT4G23700.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MGRAEEIVKMLDTAKDERTRKEALKALSAPSVSD MLSEKRRTGNKAMSIAVAGICLPFAMGTGSSFVL RATISKGVSQGQAFVGMCILTVPCRIYLFDVLSSI ALTDQFSDPERGKSELGRSQ
776	sp_090980_fpms	-0.857356	#N/A	#N/A	alpha/beta-Hydrolases superfamily protein; FUNCTIONS IN: lipase activity; INVOLVED IN: lipid metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: AB-hydrolase-associated lipase region (InterPro:IPR006693); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT1G18460.1); Has 1707 Blast hits to 1690 proteins in 186 species: Archae - 0; Bacteria - 0; Metazoa - 1166; Fungi - 276; Plants - 162; Viruses - 0; Other Eukaryotes - 103 (source: NCBI BLink).	MMQRIVDNAIAITKESVKTVTYESLSNIVRLINSISA ILLAILPGKASILEGLLHGWELRPTFRGPRLPRWM GSGVSSFNQLIHELSVDSDASSVEYSSGDEDDG YSSPASPGSQSSRVSWASSVNEYDNHRTLIGSQ HAQCPSPVRSPRKLPSLREHFVHRATDRRRGVI EDIHLATEIVIEAVFDAVHKTAHFLLSPSEAFRILS GLIFPQGNSNVHVHEHAENVPTAVLGDHEPAPT ERKTPFHPLNTDARTCQDVITEFGYPYEAIHVVTE DGYILLLERIPRRDSRKAVYLQHGVLDSSMGWVS NGVVGSPAFAAFDQGYDVFLGNFRGLVSREHVD RNISSREYWRYSINEHGTKDIPAMLEKIHEVKTSE LKLRYPDYDEDNSDEVPYNVCAICHSLGGAGILM YVITRRLEQKPHRLSRLILLSPAGFHEDSNLAFTV VEHIFLGLAPILVKIVPAFYIPTRFFRMLLNKLARDF HNYPAVGGLVQTLMSYVVGGDSSNWIGVLGLPH YNMNDMPGVSFFVALHMAQMKRTGKFRMFDYG NPDANFEVYGSPEPVDLGECFDLIDIPVDLVAGR KDKVIRPAMVRKYYNLMTEAGVEASYNEYDYAHL DFTFSHREELLAYVLTRLQLVKSTPKQRLDSVVE RSKEDLDVQSD
777	sp_091100_fifh	-0.662142	#N/A	#N/A	serine-rich protein-related; BEST Arabidopsis thaliana protein match is: serine-rich protein-related (TAIR:AT5G11090.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAATSSSRHQRSHSPSGRFCASSISSSSSSFTSS SPTFTSQSTDLFHRSSSPTRVSIYGCSSPQTTSS SVRFSIDNRPTSPNRSFAAVNRNRNPFPAASGKK KTCMCSPTTHPGSSFRCSYHKSLEAKKQQNMVSY PPNRLNMRRSAMTNSLVRIGTVEGGDSVKRALA ALIRPCAHQQRRRSDFQPRPSRLSIMSKAGDSD P
778	sp_091370_hwci	#N/A	#N/A	-0.880165	uncharacterized protein	MDQETLSCSSAYRVGDCDGNIEPQVNSSVQPYIL KKLDEEQSSLSRRTPDFNDRELPLAMEQTLSAKK KFPACINEGYGGEGVKLEDLRKADDEVVSDFSP ENDVQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
779	sp_093490_xtok	-0.88873	-0.949997	-0.890871	uncharacterized protein	MTMQRGEGWCFVVAGQRGEDCGRAGLQDGNI RILLKVDSIMQGRSYPYHDSACTCKGMPGFHFKP PWLIFFITMPPHYISHSTKIEVQDLFLLGCLLDSRT NNNHSPMSQLQTRFQARQLQRGNPSSTASGHA DGWNPQLPSL
780	sp_093850_hswi	#N/A	-0.848483	#N/A	bromo-adjacent homology (BAH) domain-containing protein; FUNCTIONS IN: DNA binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage, petal differentiation and expansion stage, E expanded cotyledon stage, D bilateral stage; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Bromo adjacent homology (BAH) domain (InterPro:IPR001025); BEST Arabidopsis thaliana protein match is: nucleic acid binding (TAIR:AT3G15605.4); Has 602 Blast hits to 478 proteins in 106 species: Archae - 0; Bacteria - 17; Metazoa - 295; Fungi - 38; Plants - 91; Viruses - 4; Other Eukaryotes - 157 (source: NCBI BLink).	MSKAEDNDERESIDFSWGTKRGRGGQNKGVQF YESFTYDGVEYFLYDSVYLSKEGEAKPYVGKLIKI WEQPNNARKVKILWFFYPREISNYLKGENVLENE LFLAKGDGKGLANVNPLESIGGKCNVVCISKDER NRQPSVEELKMAHYVFYRTFDVQHHTVTDVLGE KIAMTEVNLLLNKSDFLSLDLPMVDSANAVAGLE NPLISHTYTLVEDGTMECDATGGELLEKVNKGDT ILLDKSKYHPGGHIDAAGLDNTGLVVGRERKLVN KDPGGVDKDVVYDDSLGSLLSNDDKGSKGKSDA AVPDDLGLASVPATEVALASGSKRRLRSPESKEN DLDEKYVVLPVTDAEKKISDHVKENHRAVGAVAK VKSDDSNDKPFKKVKINGSALVSSDNGGTSTPKS AAVASNKGSQSVVVGTDSNIGGRLTKEFDGISNG HPKKLKPVEKSDDRLREMSVKQPKDKQESAATD IAANKASQPGNDDVETDDKGNLSRKVASTSNGL PKKRKSSSTIDDKSNDKLRKTFVKEPKDKEIEAY GCVKEVTERPPDRSKWFTASPWEGRMKDAHDQ GRLVLLHNLDPTYTSTEVENIVWHGFGESCTGKV VQRTATSSRCSGQAYVIFKTREVAEKVLKKLDNE CFMLPNGRPLVGSRDVVCFPGKQSQFPGHLVIP KPQMQRESREAVSTSHSAQSNTLEYEMALEWR LLQERSELAWKMLYKQQGEQIKKLKATMKSE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MGSVKRKSVEEENPSMESNPPPQKVMRENSNS
						GTVACFHDVSYPDSYVPRSSTSNSSSTSTPAKE
						FPFSLDPFQAQAIKCLENGESVMVSAHTSAGKTV
						VASYAIAMSLRNKQRVIYTSPIKALSNQKYREFKE
						EFSDVGLMTGDVTIDPNASCLVMTTEIWRSMQY
						RGSEITREVAWAIFDEVHYMRDKERGVVWEESI
						VMAPKNSRFVFLSATVPNAKEFADWVAKVHQQP
						CHIVYTDYRPTPLQHYIFPSGGKGLYLVVDEKAKF
					RNA helicase, ATP-dependent, SK12/DOB1 protein;	REDSFQKALNALAPGGGEKKRENGKWQKGVVV
					FUNCTIONS IN: hydrolase activity, acting on acid	GKLGEESDIFKLVKMLIQRQYDPVILFSFSKRECE
					anhydrides, in phosphorus-containing anhydrides,	FLAMQMAKMDLNGDDEKVNIETIFWSAMDMLSD
					helicase activity, ATP binding, ATP-dependent	DDKKLPQVTNLLPLLKRGIGVHHSGLLPILKEVIEI
					helicase activity, nucleic acid binding; INVOLVED IN:	LFQEGLIKCLFATETFSIGLNMPAKTVVFTNVRKF
					N-terminal protein myristoylation; EXPRESSED IN: 23	DGDKFRWISSGEYIQMSGRAGRRGIDERGICILM
					plant structures; EXPRESSED DURING: 14 growth	VDEKLEPSTAKMMLKGNADSLNSAFHLSYNMILN
					stages; CONTAINS InterPro DOMAIN/s: DNA/RNA	QLRCEDGNPENLLRNSFYQFQADRAIPDLERHA
					helicase, DEAD/DEAH box type, N-terminal	KQLKEEKDSIIIEEEETLKNYYSLIEQYKSLKKDVR
					(InterPro:IPR011545), DSH, C-terminal	VIVYSPKYCLPFLQPGRLVCIECTKTDEDPPSFSI
					(InterPro:IPR012961), DEAD-like helicase, N-terminal	EDNVTWGVIINFQRIKSTSEDESSKKPEDASYTL
					(InterPro:IPR014001), DNA/RNA helicase, C-terminal	DVLAKGMVSREGGAKKTTKIVPLKDPGVPVVVSV
					(InterPro:IPR001650), RNA helicase, ATP-depencent,	PLSQVYNLSCIRVIMSKDLLPLEARENTLKKVSEVI
					SK12/DOB1 (InterPro:IPR016438), Helicase,	SRCSREGVPLLDPEADMKVQSSSYRKAVRRIEAL
					superfamily 1/2, ATP-binding domain	ESLFDKHEVAKSPLIKEKLQTLHKKQELAAKIKSIK
					(InterPro:IPR014021); BEST Arabidopsis thaliana	RQMRASTALAFKDELKARKRVLRRLGYVTNDDV
					protein match is: RNA helicase, ATP-dependent,	VELKGKIASEISSADELTLTELILNGVFKDIKVEDM
					SK12/DOB1 protein (TAIR:AT2G06990.1); Has 9909	VALLSCLVWQEKLQDAQKPREELELLFAQLQDT
					Blast hits to 8438 proteins in 1290 species: Archae -	ARTVAKVQLECKVQIDVENFVSSFRPDIMEAVYA
					828; Bacteria - 3072; Metazoa - 1170; Fungi - 1231;	WAKGSKFYEIMEITQVFEGSLIRAIRRLEEVLQQLI
					Plants - 476; Viruses - 32; Other Eukaryotes - 3100	LAAKSIGETELESKFEDAVSKIKRDIVFAASLYL
781	sp_094100_cqxz	-0.3113421	-0.8671149	#N/A	(source: NCBI BLink).	
						MKLKINKACDLNSISVLPPHTRRSSMMQQPRSQP
						SQQSFSQGFSSQQNGIFSQISQNSIDVIHADNQR
						LSSQDKENSVKRLSYLAPASYPREESQMPISRTS
						SSLTRRWNSSTGPENKGQLSDELEHKIAMMDTS
						LNKLGMILDSVQADILQVNKGTKELSLEVEGIRQK
						SICHHDSLHLLNRTQEETKATLESISNQLLTIQKD
					putative recombination initiation defects 3 (PRD3);	KSQEILLAVLKLQEQNDINYQNIQKHNDELPKYIS
					Has 279 Blast hits to 273 proteins in 87 species:	GVTRDLKECEHSRFKEIMCSIQTLHKSAPLTILRP
					Archae - 2; Bacteria - 106; Metazoa - 105; Fungi - 5;	KVSAAGSSLPKPFTMREAEPTKVSKPAHLVQKIEI
					Plants - 32; Viruses - 0; Other Eukaryotes - 29	GRWTTVKAEETAFSNQHGQNPKQRRPPLDKKQ
782	sp_094760_mpct	#N/A	-0.8663327	#N/A	(source: NCBI BLink).	KVIIDLDNETDEDLSCLIVEKKSDNNFMDELEDT
						AEILRKARRRKRKQDCKRSFLTHRQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
783	sp_095490_okqa	0.884823	#N/A	#N/A	Protein prenyltransferase superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: Tetratricopeptide TPR-1 (InterPro:IPR001440), Tetratricopeptide-like helical (InterPro:IPR011990), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734); BEST Arabidopsis thaliana protein match is: Protein prenyltransferase superfamily protein (TAIR:AT5G37130.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAEIEVQHLLEYELRLLRCTLPQPPSNTRRFLQP HDTPSHPPFHLLIDNIVSLIESGRYVDSLSSAGVR TVFQFSNSFESLDPSQFYAGLEKAVESFLVDDSA EGHASFKFFLVVSIAIAAILAFVQCNFTGPVVELPS VPLPWKVIKEEKEWEAWARIQLMSDGSDLAAKF SNLQYILFAKMLLMKTRDLSFGEKTVCEAGLGTS TWWFARVIFLQQRILDECCFSLLNLLQVFMSESL NRFGNVEKVTSYWGASLAEDEALTIVSMLQLEA GLVEHTFGRVDHARQYFEAAAFSVGLELSVTGA LGVRTVHQVDPVAQSVLVARKSSFPSASMSPLE NASHVRDASHDPNSHQNHQPYSDTSDVLMAPK FMDSGGNDLNLIEHGCATDRTLTAIQQAVVLAHC LLIEKRARFDEQQSWDMAPYIEAIDSQLSLYFTIK CSCNILRIKWESTRSRTKERALLMMNKMVENLYE PSPQVLERVHSCFSVLIPAAPILRKEYGELLISSGL IAEALGIFEDLELWDNLIFCYRLLEKKAAAAELIKV RLLETPSDPRLWCSLGGVMNDDSCYEKALEVSN DRCARAKRSLARSAYNRGDYEKSKVLWDSAMA LNALYPDGWFALGAAALKARDVSKALDGFNRAV QLDPDNGEAWNNIACIHKIRKKSQEAFIAFKEALK SKRNSWQMWENYGQVAADVGNYFEALKAAQM VLDITGNKRYDGDLLERIVQEMEERFSSYSPITED DKDCNSQTPSGSELAESRDTKLLMELLGKVLQQ VVRSVGSGDSWGMYARWHKMRGDLTMCSEAL LKQVRSYQGSEVWKDNARFRKYANASLELGKIY MVISSSSESLRELHSADMHLRNTIKQAISFTDTEE FKALQVCLDDVQMRLKSSLVPTS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
784	sp_095600_pxrk	-0.873003	#N/A	#N/A	SIN3-like 4 (SNL4); INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Histone deacetylase interacting (InterPro:IPR013194), Paired amphipathic helix (InterPro:IPR003822); BEST Arabidopsis thaliana protein match is: SIN3-like 3 (TAIR:AT1G24190.2); Has 2571 Blast hits to 1108 proteins in 232 species: Archae - 4; Bacteria - 56; Metazoa - 1028; Fungi - 808; Plants - 415; Viruses - 14; Other Eukaryotes - 246 (source: NCBI BLink).	MKRSRDDVFTGPQVKRPQPANSYRGESSGQPQ MVGGGGVGAPTPQRLTTNDALAYLKAVKDMFQ DKKEKYDEFLEVMKDFKAQRIDTTGVIARVKDLF KGHRNLILGFNTFLPKGYEITLPPEDDNPPAKKPV EFEEAISFVNKIKLRFHGDERVYKSFLDILNMYRK ENKTISEVYQEVAALFRDHRDLLEEFTHFLPDTSA ATPGQNGPSGKNSILRDRNSGMPSTRLLHVDRK DHEFNLDGSDADNDKLGKAERDQRRRLDKERD RKDGINRVEDSITDHGAENYGPRAVSGSYDDKS GLKSMYSQELAFCEKVKEGLDNLDDYQEFLKCIH IYCREIITRQELQALVRDILANYQDLVDGFDEFLAR CERPDGFLSNVMSKKSLWPEFLCSDGQSKSVKT EDRERDRDMVRDDRDRDRESRERDRTDRGVIY GNKEVANLRLSSNSNKDKYIGKPIQELDLSNCER CTPSYRLLPKNYPIPLASHRTELGAGVLNDHWVS VTSGSEDYSFKHMRKNQYEESLFRCEDDRFELD MLLESVNVTIKRVEDLLDKINSNMIKTENPIRIEDH FTPINLRCIERLYGDHGLDVMDVLRKNAILSLPVIL TRLKQKQEEWARCRADFNKVWAEIYSKNYHKSL DHRSFYFKQQDTKSLSTKALLAEVKEISEQKCKE DDVLLAVAAGNRRPIIPHMEFEYPDTELHEDLYQL MKYSCGEVCTAEQLEKVMKIWTAFLEPVLGVPA RAQGAEDSEDAIRAKNEASKTDGPATGESDRSP SGGAATRKHESSPPEQSSSGNHAKDEGYPDAS SSGLHQVRSQSNVIDEFGLGKQTTSNERLTASN PSLPSGVEQFHGRNIENPSGHNTTQLRVGNAAD NSGIELRSNGGNIPSLEAVDSTRPVAASNGVITAA TKEIRYCEESNGHAKVEREEGELSPNGDFEEDN FAVFGDASTEPANKPNDGASSAPHRSGPRGARR ENDVDADDEGEESAHRSSEDSENASENGDVSH SESGEGEDCSPENEEDGEHDENDNKVESEGEA EGMADAQDGDGISLPLSERYLQTVKPLAKYVPAS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
785	sp_096400_wtuh	#N/A	-0.885782	#N/A	polypyrimidine tract-binding protein 2 (PTB2); FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: pollen germination, regulation of RNA splicing, regulation of translation; LOCATED IN: nucleus, cytoplasmic mRNA processing body, cytoplasm; EXPRESSED IN: 31 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: polypyrimidine tract-binding protein 1 (TAIR:AT3G01150.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MASVSSQPQFRYTQPPSMVLHLRNLPWECTKE ELIELRNPFGKVVNTKCNVGANKIQTFIKFADLSL SIAMISYYASSSVPAQIRG
786	sp_096760_dcxs	#N/A	#N/A	-0.8999553	DDT domain superfamily; CONTAINS InterPro DOMAIN/s: DDT domain (InterPro:IPR004022); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G25580.1); Has 20293 Blast hits to 13585 proteins in 823 species: Archae - 41; Bacteria - 1919; Metazoa - 8043; Fungi - 2539; Plants - 1031; Viruses - 220; Other Eukaryotes - 6500 (source: NCBI BLink).	MQSYDLLPNPTLISPNLHLLDLPVMEHVVSSPIHR SLGEPETNGDAPKPESNNYPETGPEAGNKEGSN PIEEESKKELSGNDVGSKSSSKNGKKENNSSSN NNVIVRRERPTRACTARPAKYVDPPVIERRPRPP KRERVEKVVAEEVSEEEDDDLPKEQCSKVVTSL VNPPTQEELMPRWELRSMWELASILNFLHIFRAV LNINAEFTVEELESALLSPNSTLSDIHIPLLKAIPPV TRMACGRDTWVTVLCRKLKDWWHWVAEGECP VVTDHGAETEAYKNLDPKVRVVILKALCDIRVEQ EDIRSYIDSSLKHGAQLSAFRKERIGGDSHGTSF WYEDDEFVGHRLYREIRKVEKKVEFKKGKTKGS NVLPCTTYEWETVATNLDEFLDVSEKLFTSKNRT EASLGKKLRNDMLPEIEKIHKRKEKMLKKQQRQA LLLDNFVMDGLAGGRSLRDRKPVTYTFDDFDRSI NEAIKTTKRKQPSPEPGVKREVMKPEPSTNGRW NGSSNSSLHVSFSPPSPVDYDEMDVDHNSEPLD RGNRRRQRPQRYSEKEFVEAVSDNDEEYFYSD DDIVGEAVYDEAYMQQRKQKKRQTSSSEGEGD EEYQWSDENGVDEEEEEEDDSVSNNSEDSDQP RGRFKKPEGRTRSRRGTKIRSVDDIDDSGLRRS NRATRNRIDYRTYELSDSETETSTKKGKANLHAL DNAGYSSESQDSDEGNEEEEEEQQEKGVGSES APMAVENYDNEPEPEPLAKSESPKQNDPERRIE ATRGYLDLNELAPGSGFDDGPGALLKDDDDDQK GE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
787	sp_097120_jway	-0.904084	-0.907516	-0.91591	KAKTUS (KAK); FUNCTIONS IN: ubiquitin-protein ligase activity; INVOLVED IN: trichome branching, DNA endoreduplication; LOCATED IN: plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-like helical (InterPro:IPR011989), Armadillo (InterPro:IPR000225), Armadillo-type fold (InterPro:IPR016024), HECT (InterPro:IPR000569); BEST Arabidopsis thaliana protein match is: ubiquitin-protein ligase 4 (TAIR:AT5G02880.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	METRSRKRAEASPSAPSSSSAGPTTRAAKRTRL SSSSSSTTAATPATRSRTRSNISVSPATMDESSS RGGRRGRNHHGADKDNDKGKEKEHEVEKEREN ELGGLDVFDDDDDDNDNDDSEGGGGVGLLHQN LTSASSALQGLLRKLGAGLDDLLPSSAMGSASSS SHQSGRLKKILSGLRADGEDGKQVEALTQLCELL SIGTEDSLSTFSVDSFVPVLVGLLNYESNPDIMLL AARAITHLVDVLPSSCAAVVHYGAVNCFCARLLTI EYMDLAEQSLQALKKISQEHPTACLRAGALMAVL SYLDFFSTGVQRVALATAANMCKKLSSDAADFV MEAVPHLTNLLQYHDAKVLEHASVCLTRIVEAFA SSADKLDELCNHGLVTQVASLISTSNSGGGQASL STPTYTGLVRLLATCASGSTVGAKTLLLLGVSGIL KDILAGSGLVGNMSVSPALSRPPEQIFEIVNLANE LLPPLPQGTISLPASSNLFVRGPLFKKSPMSSSG KQDDNGNVQDVSAREKLLNDQPELLQQFGIDLL PVLIQIHGSSVNGPVRHKCLSVIGKLMFFSNADMI SSLNNTTNISSFLSGVLAWKDPQVLVPALQIAEIL MEKLPETFSKMFVREGVVHAVDSLIASPSNSVSA HTATPEKDTESTLGPSARARRLRRRSGNHQQDA EELKNSVSVGGSPPTSVEVPTASSNLRTTVSGC AKSFKDKYFPSDPGTVDVGISDDLLHLKNLCAKL NAGVEEQKSKVKGKSKTAGAYVAEFVTGKEESL VGIVSEMLGELSKGDGVSTFEFIGSGVVAALLNY FSCGYFNKERILEVNLPKLRQQALKRYKLFISIALP NEGNVVPMTILVQKLQNALSSLERFPVVLSHSAR SSGGSARLSSGLSALSQPFKLRLCRAAGEKSLR DYSNTVVLIDPLASLAAVEDFLWPRVQRTESGQK ASTSVGNSESGPLPTGAGASSPAASTPATSTRR QSTRARSGVNIGDTSRKDASQEKNNSSSKGKGK AILKPGTEEIKGPQTRNAARRRAVSDKDAQLKSP TEESSSEDEDLEMSPVDIDDALVIEDDDISDDEDD
788	sp_097150_jzay	-0.869399	#N/A	#N/A	uncharacterized protein	MLKYAYGKYLGLVDYLHRTGLAKKCAGLRMGAK VRAFPGFVLGDL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
789	sp_097220_oxuk	-0.881794	#N/A	-0.91953	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: leaf senescence, cell death, ethylene biosynthetic process, defense response; LOCATED IN: nucleolus, nucleus; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT2G41060.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MDASKKRKADENGNSFLSDGTLTSPSDLSPPAQ AQTLTPDDARKLIEPFTIDQLRTLLQDAVVRHPEV LVSTRALADSDVSRRKLFVRGLGPTTTDSLKSLF TSYGELDEAVVISDKTTGKSKGYGFVTFKHVDGA LLALRQPSKKIDGRMTVTHLASAGRGPNSEDQA TRKIYVGNVPYEISAERLLDYFAAFGEVEEGPLGF DKQGGKARGFAFFVYKTEEGARMSLMEPMKMID GHQVMCKLAEDGKKGKMGGPNPVMPGEMGGD RMSGSMNPQTGSYGNYGPGYNSGPGMGMGAA QGGGGGPYGLHSQYPGSGEYASYRMPPPQSS MGMPPPYQEGSGYGPPSSGGGGGGYGPPSSG GGGGYMGQNQPPSQGPPQGPPGGMYQRMPP SYY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
790	sp_098670_qwid	-0.856864	#N/A	#N/A	actin binding; FUNCTIONS IN: actin binding; INVOLVED IN: cellular component organization, actin cytoskeleton organization; EXPRESSED IN: shoot apex, stem; CONTAINS InterPro DOMAIN/s: Actin-binding FH2/DRF autoregulatory (InterPro:IPR003104), C2 calcium/lipid-binding domain, CaLB (InterPro:IPR008973), Tensin phosphatase, C2 domain (InterPro:IPR014020), Actin-binding FH2 (InterPro:IPR015425); BEST Arabidopsis thaliana protein match is: Actin-binding FH2 protein (TAIR:AT5G07770.1).	MALFRRLFYRKPPDRLLEISERVFVFDCCFSTDAL DEDEYTEYLGGIVTQLQDYCPDASFMVFNFREG DRRSHIADILSDYDMTVMEYPRQYEKCPLLPLEM VHHFLRSSESWLSLEGQQNVLLLHCERGGWPV LAFMLAGLLLYRKNYSGEQKTLEMVNKQASKELL QLLSPLNPQPSQQRYLQYISRRSLGSDWPPSDT PLILDCLILRVLPLFDEGKGCRPVVRVYGQDPLTK SSRSTKQLFSSLKSKKQVPRYSETECLLVKLDIR CRVQGDVVLECVHLCDDLIREEIMFRIMFHSSFIR SNILSLSRDEVDKLWDAKVQISKDFKAEVLFSDV GAMPIINTELESDDDDNETDSASPEEFFEAEEIFS NAVDAQDGKTDFDGPFIHDSSVNEVGNGEQKED REPPSFVDCGPDTRAHRRDIKELVEIETVKDIGLD GWSKTLHRHVESDLNAIKDILMDDANQKEDVKW AVDDIVLHEGGHKHEVHYDLNEVKDIALDAYVKE ADTMIVVSDTLIKQNIKTGFVDGNMVGELDRKQT VEDWEMGKGSDSSNSPRKKLNGDVTKQKQERV LQPSPKRPPTFNQKSASDSALGKHRGKQQEAAG DKSGKQKPTEHRWIPSNKGSYTNSMHVAYPPTR VNSAPPSLGGTKNQVKDSGSVGKLKDRAASTDF GFGATSIDRRPHKFQDVNPVESAKFSWVETPDD GLNQEVSSPKCFPRSDPSSLVNPSSQSLAESVE LEEEIKAPLTLPPQSPVHSSSSDMPTVEVLPLQR PPSPSSINSCGNFKISSPMLPQSPSLFDSSDKQS PPPPPPPPGGRGPPPPPPPPGGRGPPPPPPPP GGRGPPPPPPPPGGRGAPPPPPPPGGRGAPPP PPPPGGRGPPPPPPPLRGGGPPPPPPPRGGAP GPPAPPRPPGAGPPPPPPFGAKAAPGAPNAPPS LGRGRGLPRSAPGVATKKTTLKPYHWIKVPRLIQ GSLWEELQTQAGSQIVAEFDMSEIEKLFSNQVSK PAGAGGKSGEKKKAVGSKSEVVHLIDTRRVNNT EIMLTKVKMPLPDMMAAVLAMDESILDADQVENL
791	sp_098770_mxac	-0.941301	-0.919205	-0.959234	RAN GTPase 3 (RAN3); FUNCTIONS IN: protein binding, GTP binding, GTPase activity; INVOLVED IN: protein import into nucleus; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Ran GTPase (InterPro:IPR002041), Small GTP-binding protein (InterPro:IPR005225), Ras (InterPro:IPR013753); BEST Arabidopsis thaliana protein match is: RAS-related nuclear protein-1 (TAIR:AT5G20010.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MALPNQQTVDYPSFKLVLVGDGGTGKTTFVKRH LSGEFEKKYEPTIGVEVHPLDFHTNCGQIRFYCW DTAGQEKFGGLRDGYYIHGQCAIIMFDVTARLTY KNVPTWHRDLCRVCENIPIVLCGNKVDVKNRQV KAKQVTFHRKKNLQYYEISAKSNYNFEKPFLYLA RKLAGDQNLHFVESPALAPPEVQIDLAEQQRHE QELAVAAAQPLPDDDDDAFE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
792	sp_099050_hcsn	#N/A	-0.8655885	#N/A	Major facilitator superfamily protein; FUNCTIONS IN: transporter activity; INVOLVED IN: oligopeptide transport, pollen tube growth; LOCATED IN: membrane; EXPRESSED IN: petal, leaf whorl, sepal, male gametophyte, flower; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Oligopeptide transporter (InterPro:IPR000109), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: nitrate transporter 1.7 (TAIR:AT1G69870.1); Has 3642 Blast hits to 3525 proteins in 525 species: Archae - 0; Bacteria - 558; Metazoa - 480; Fungi - 335; Plants - 2158; Viruses - 0; Other Eukaryotes - 111 (source: NCBI BLink).	MGILTVSVGVPKLKPPKCEPPAECLPSESWQLAI LYAGLTLVAIGGGCVRPCNIGFGADQFDVTTEKG RASYARFYNWWYFSFTFALLVALIGVVWIQVHIS WAIACVVATVCFASSTIIFLIGSPTFINKKPQGSIYI DMARVVVAAFGKRNVQNHQTIYDPPCISNSNDVI APPKLSRTNRLIYLDKASLITQPEELNNSGLAINK WSLCSVGQVEQLKCLFGILPVWLTGVFCFVSIDQ HNSYGILQAMQMNRRFTRKGFEIPAAWIGFASM ASLSLWILAYERICLPLAKKTKKNDVRLSINTRIKI GIVMSVLSTSTAAVVECKRRELAIKNNSYSSPLHV GYLAPQQVICGLTEAFAGVTVMEFYTTQMPESM RSLSGSIFFLNLAMASYISIAINIITKITQHTHQGS WLGGNDLNHNRLDYLFAIVACLGVVNYFFFTTFA SKFVFIDNNIETTEAASSSTEVHPRVE
793	sp_099410_ujgo	#N/A	-0.8735317	#N/A	NADPH-dependent thioredoxin reductase A (NTRA); FUNCTIONS IN: thioredoxin-disulfide reductase activity; INVOLVED IN: in 6 processes; LOCATED IN: cytosol, mitochondrial matrix; EXPRESSED IN: male gametophyte, guard cell, cultured cell, pollen tube, leaf; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, seedling growth, seed development stages; CONTAINS InterPro DOMAIN/s: Pyridine nucleotide-disulphide oxidoreductase, class-II, active site (InterPro:IPR008255), FAD-dependent pyridine nucleotid-disulphide oxidoreductase (InterPro:IPR013027), Pyridine nucleotide-disulphide oxidoreductase, class-II (InterPro:IPR000103), Pyridine nucleotide-disulphide oxidoreductase, NAD-binding region (InterPro:IPR001327), Thioredoxin reductase (InterPro:IPR005982); BEST Arabidopsis thaliana protein match is: NADPH-dependent thioredoxin reductase B (TAIR:AT4G35460.1); Has 20135 Blast hits to 20133 proteins in 2901 species: Archae - 588; Bacteria - 13597; Metazoa - 140; Fungi - 273; Plants - 163; Viruses - 0; Other Eukaryotes - 5374 (source: NCBI BLink).	MSSSSPPNAQMSRCLKLRTLLKRSLNFLPPLSLS SSSSPAAAATTTTSLTSSAANSLSTVAAPTMEDF KQALRTRLCIIGSGPAAHTAAIYASRAELKPILFEG WMANGIAPGGQLTTTDVENFPGFPEGIFGLDIT DKFRQQSLRFGTEILTETVTSVDFSTTPFKILSAE SKTVLADSVIIATGAVAKKLEFPGSGEEGGFWNK GISACAVCDGAAPIFRNKPLAVIGGGDSAMEEAN FLTKYGSKVYIIHRRGEFRASKIMQKRAIENPKIEV LWNSTVTEAYGDAEKENRVLGGLKVKDVVSGEV KDLQVSGLFFAIGHEPATKFLGGQLELDADGYVV TTPGTTKTSVKGVFAAGDVQDKKYRQAVTAAGT GCMAALDAEHYLQEIGSQEGKSD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
794	sp_100510_fgkr	#N/A	-0.873829	#N/A	co-factor for nitrate, reductase and xanthine dehydrogenase 7 (CNX7); CONTAINS InterPro DOMAIN/s: ThiamineS (InterPro:IPR003749), Molybdopterin converting factor, subunit 1 (InterPro:IPR010034), Molybdopterin synthase/thiamin biosynthesis sulphur carrier, beta-grasp (InterPro:IPR016155), Beta-grasp fold, ferredoxin-type (InterPro:IPR012675); Has 1000 Blast hits to 1000 proteins in 381 species: Archae - 21; Bacteria - 814; Metazoa - 52; Fungi - 2; Plants - 54; Viruses - 0; Other Eukaryotes - 57 (source: NCBI BLink).	MEQKNQIESQGNNTHERNEKSSIQLKVLFFARAR DLTGLADMPLEVTSGSTARDCLDKIVTQFPGLEEI RGCMVLALNEEYASESAIVQHRDELAIIPPISGG
795	sp_100590_qojs	-0.936395	-0.966726	-0.903907	Ribosomal protein L10 family protein; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation, ribosome biogenesis; LOCATED IN: cytosolic ribosome, ribosome, chloroplast, chloroplast stroma; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal protein L10 (InterPro:IPR001790); BEST Arabidopsis thaliana protein match is: Ribosomal protein L10 family protein (TAIR:AT3G12370.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGATISTTTTSLSHHTLTLIRTQSHHHHQLRHFAK YTPYKPLPKPKVHQKLRAIKTLKSHLQDSYLISPF PCHGFSANKFKPLKTPLPSDPSAAKLVFVKNSLV SQAFAGTQFEALGSCMKGMNSLLFVKKEDAVGP VMGGVKSAVKEGKLEFNDFPGAVVNGQLYGHLD LEVFQNTPSKMEAHGMVLGSLFGPSSTLMAVLE GFCADDSANEAATDAPADGGIDAKDVSESPAQ
796	sp_102600_rozo	#N/A	-0.869894	#N/A	uncharacterized protein	MGTSDILPEKADTHGCLYCCSRDIHHPADICSFR SGGEPHFDTLCSKLKEFNKYLEKGILRKRKDKED LSLRFPDLGLEYRGNNTCTDMIARNASVALRLFN GEKGSNFELVKIGRFEIVVLGFGFLQHMNFEAKN KDDPDAEVMTFFAEVIRRARYFFVDFCQCLGPSF LLPDEAELKGCCYCKCIHHPPGPGYLIGRELD
797	sp_103080_jhua	-0.885664	-0.889073	-0.858549	unknown protein; Has 19 Blast hits to 19 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 19; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MEVGSKRENEKQKKTEKPPYVVAKDDTKPVLQD PLWSSDPTQAEQAVLRLPPFPYKSSNDQLN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
798	sp_103380_mzes	-0.847734	#N/A	#N/A	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: membrane; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage, F mature embryo stage, petal differentiation and expansion stage, E expanded cotyledon stage; CONTAINS InterPro DOMAIN/s: Aminotransferase-like, plant mobile domain (InterPro:IPR019557), Protein of unknown function DUF716 (InterPro:IPR006904); BEST Arabidopsis thaliana protein match is: Aminotransferase-like, plant mobile domain family protein (TAIR:AT1G51538.1); Has 16736 Blast hits to 9656 proteins in 576 species: Archae - 4; Bacteria - 1182; Metazoa - 7098; Fungi - 2631; Plants - 1178; Viruses - 174; Other Eukaryotes - 4469 (source: NCBI BLink).	MGSFKGHVVPGTLFLLVGLWHMWCSVYRVRS PNTFRIRVWNPVPGFDGRIKYLELYVVVIGSFIDL CIELVYSPHLKYFVNGVLNPSHMNNFEHSGMLIM FFIFGLVGLLSVKTSILPLPDGALTLIGATAFSSEY LLFYFHSTNHKGLEGYYHYILVLLIGLCVLTSVVG TLVPSNFPLDLSNSIAITLQGLWFYQTAFTLYGPM MPHGCQLKGDDIVCHSADSQIRGELLANFQLFS MVVGVFAVISVFYSFATSRFGHSSQERLQEADG DL
799	sp_103510_diku	#N/A	-0.837155	-0.85312	MUTM homolog-1 (MMH-1); CONTAINS InterPro DOMAIN/s: Ribosomal protein S13-like, H2TH (InterPro:IPR010979), DNA glycosylase/AP lyase, H2TH DNA-binding (InterPro:IPR015886), DNA glycosylase/AP lyase, catalytic domain (InterPro:IPR012319), DNA glycosylase/AP lyase (InterPro:IPR000191); Has 11102 Blast hits to 10156 proteins in 2240 species: Archae - 10; Bacteria - 5065; Metazoa - 1812; Fungi - 570; Plants - 321; Viruses - 42; Other Eukaryotes - 3282 (source: NCBI BLink).	MPELPEVEAARSAIAEHCIGKKILRSITADDSKVIN GVSRSHFEAALLNKTIVAAHRKGKNMWLQLDCP PFPTFQFGMAGAVYIKGVAVTKYKRSAVSDNDE WPSKYSKFFIELNDGLEMSFTDKRRFATVRLLDD PASAPPISALGPDALLEPMTEGEFNKSLSKKKIAI KPLLLDQSFISGIGNWVADEVLYQARIHPSQNAS SLSKDESDRLLKSITEVLEKAVEVGADSSQYPSG WIFHSREKKPGKAFVDGKKIDFITLSGVQATEAP AKPRKRASKKKDTDDHNDDGDDTNGDVREEETE DTKSKQTKTSKSGKQPARRKKSTKSNDNDDDDD DVDGGGGGGGEDDGDGDDKTTGTRGGKVIAK RKLIGTNESQDANKSKKTRGGGKKDTAEPKSRK SNDSDASDDDGNENGDQKAGAKKPQKETRKKA K
800	sp_103740_qrrz	#N/A	-0.837011	#N/A	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity; INVOLVED IN: oxidation reduction; LOCATED IN: nucleus; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Taurine catabolism dioxygenase TauD/TfdA (InterPro:IPR003819); Has 1029 Blast hits to 1021 proteins in 229 species: Archae - 0; Bacteria - 729; Metazoa - 46; Fungi - 17; Plants - 110; Viruses - 0; Other Eukaryotes - 127 (source: NCBI BLink).	MAESFTGIQTPEQKHYNSNNSFPFPLILTPNSSPL SNYNFIESIKTHKNWLNSRLHSAGAILFREFPVST AAEFNDVVEAFGYEEFPYIGGAAPRSNVVGRVFT ANESPPDQKIPFHHEMAQVPEFPSKLFFYCEVEP KNGGETPIVLSHIIYERMKEKHPEFVRELEKHGLM YTRVLGEGDNPESPIGRGWKSTFLTDDKSIAELR AAKLGMKLEWLEVGVKTVMGPIPAIKFDESRQH NVWFNSMVAAYTGWEDDLNDPVKAVTFGDGKP LPADIVYDCLKILEEECVAIPWKKGDVLLIDNWAV LHSRRSFDPPRRVLASLCK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
801	sp_104730_tqqk	#N/A	#N/A	-0.8593859	glutamate dehydrogenase 3 (GDH3); FUNCTIONS IN: oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor, oxidoreductase activity, binding, catalytic activity; INVOLVED IN: regulation of nitrogen compound metabolic process; LOCATED IN: mitochondrion; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Glutamate/phenylalanine/leucine/valine dehydrogenase (InterPro:IPR006095), Glutamate/phenylalanine/leucine/valine dehydrogenase, C-terminal (InterPro:IPR006096), Glutamate dehydrogenase (InterPro:IPR014362), NAD(P)-binding domain (InterPro:IPR016040), Glutamate/phenylalanine/leucine/valine dehydrogenase, dimerisation domain (InterPro:IPR006097); BEST Arabidopsis thaliana protein match is: glutamate dehydrogenase 1 (TAIR:AT5G18170.1); Has 8937 Blast hits to 8931 proteins in 2306 species: Archae - 305; Bacteria - 5351; Metazoa - 356; Fungi - 208; Plants - 399; Viruses - 0; Other Eukaryotes - 2318 (source: NCBI BLink).	MNALAATNRNFKLASRILGLDAKLEKSLLIPFREIK VECTIPKDDGTLASFIGFRVQHDNSRGPMKGGIR YHPEVDPDEVNALAQLMTWKTAVANIPYGGAKG GIGGNPSDLSISELERLTRVFTQRIHDLIGIHTDVP APDMGTNSQTMAWILDEYSKFHGHSPAVVTGKP IDLGGSLGRDAATGRGVLFATEALLNESGKTISG QRFVIQGFGNVGSWAARLISELGGKVVAVSDISG AIKNKNGLDIDSLLKHVQENRGVKGFHGSDAIDS DSILVEDCDVLIPAALGGVINRENANEIKAKYIVEA ANHPTDPEADEILKKNGVVILPDIYANSGGVTVSY FEWVQNIQGFMWEEERVNNELKTYMTKGLKNV KEMCKTHNCDLRMGAFTLGLNRVARATVLRGW EA
802	sp_104780_rkym	#N/A	#N/A	-0.8704446	PANB2; CONTAINS InterPro DOMAIN/s: Pyruvate/Phosphoenolpyruvate kinase, catalytic core (InterPro:IPR015813), Ketopantoate hydroxymethyltransferase (InterPro:IPR003700); BEST Arabidopsis thaliana protein match is: ketopantoate hydroxymethyltransferase 1 (TAIR:AT2G46110.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MSTHVISSSILLHISLPSLRSSFNSPGAEFGTVKV VETALALQEAGSFAFVLECVPAPVAAATTFVLGIP TIGIGAGPFCSGQAIDSSVRILKEGGMDAIKLEGA SPSRITAAKAIVEARIAVTPQAISVLGGFSRPQGR NISTAVKVAELVVPKAAMCSEKYNQTVKVGAER GFRVTNYMTLVPVKVFEDQGGQTHTIAAEVNGA DVPVAAAY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
803	sp_105380_osku	#N/A	-0.653331	#N/A	phosphoenolpyruvate carboxylase kinase 1 (PPCK1); FUNCTIONS IN: protein serine/threonine kinase activity, kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Calcium-dependent protein kinase (InterPro:IPR020642), Calcium/calmodulin-dependent protein kinase-like (InterPro:IPR020636), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: phosphoenolpyruvate carboxylase kinase 2 (TAIR:AT3G04530.1); Has 121330 Blast hits to 119354 proteins in 3292 species: Archae - 175; Bacteria - 14755; Metazoa - 43495; Fungi - 13131; Plants - 28233; Viruses - 522; Other Eukaryotes - 21019 (source: NCBI BLink).	MYDSFKNEYQLCEEIGRGRFGIVYRCFSPSSLAS FAVKSIDKRHLLDDDLDRQCLDKEPKILQILSPHP NILQIHNIFDSDSRLLIVTDLCKESTLYDRIVSSGPF TEPDACSIFTQLAEALHHCHRNYICHRDIKPDNILF DTRNRVKLCDFGSAEFFGCGGEREKMSGVVGT PYYVAPEVLAGREYNEKVDVWSAGVILYIMLGGV PPFYGDSVEDTFAAVLRGNLRFPTKIFRSVSPEA KDLMRKMLCKDVSRRFSAEQVLRHPWVTSGGL SNTMSDMS
804	sp_105590_iumr	-0.962653	-0.975161	-0.953087	pre-mRNA-processing protein 40C (PRP40C); CONTAINS InterPro DOMAIN/s: FF domain (InterPro:IPR002713), WW/Rsp5/WWP (InterPro:IPR001202); BEST Arabidopsis thaliana protein match is: pre-mRNA-processing protein 40A (TAIR:AT1G44910.2); Has 7513 Blast hits to 5748 proteins in 436 species: Archae - 33; Bacteria - 336; Metazoa - 3654; Fungi - 967; Plants - 481; Viruses - 7; Other Eukaryotes - 2035 (source: NCBI BLink).	MFSSSRSLQNRLYVIRIDSKVSFSCLCTYLLQFGD MLKERGVAPFSKWDKELPKIVFDPRFKEIPTYAE RRSLFEHYVRTRAEAERKEKPAAQKAAVEGFKQ LLEEAKEDIDHTTTYHTFRKRWGHDQRFEALERK DREVFLNERVLSLKRVAEEKIRAERTVVVSNFKS MLRERGDVTASSRWSRVKDGLRDDSRYKAVKH EEREALFNAYISELKAADDEVEREAKAKREEQEK LKERERELRKRKEREEQQMERVRSKVRRKEAVA SYLALLVETIKSHQASWTESKPKLVKDPQGRASN PDLDESDLERLFREHVKMLYERCACDFRALLVEK ITPEAATRETEDGKTVLNSWSTAKKFLKSDPRYT KMPRKDREALWKRHAEELQRKQKGAVEQKKGDQ YNESKNRNANDSGRPLSVSRSGQDWR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
805	sp_105830_yqri	0.827577	0.834248	0.980635	D6 protein kinase like 2 (D6PKL2); FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: nucleolus, nucleus, cytoplasm; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: D6 protein kinase like 1 (TAIR:AT4G26610.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MKSQNFACRVVSPSGESSSCNQVSGFESKCSIK
806	sp_105910_hknn	0.8663386	#N/A	0.911138	unknown protein; Has 30 Blast hits to 30 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 30; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	YPVKDSRPMIQEAHWVTQLSVYRGQRHRQSVT EASECHSIALRNRVKVERTINPFDFVERRKVSSTY RCSTGSDVNDVRWIAIRKAQSNYGLLQFTHFRLL NHLGCGDVGKVYLSEFVGTKTHFAMKVMNKAAL ESRKKLLRAQTEREILQSLDHPFLPTLYTHFETEN YSCLVMEFCPGGGLHSLRQRQPHKGFTEHAAR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
807	sp_106130_syqe	#N/A	0.865238	#N/A	uncharacterized protein	MALMREQLRARLASSKTSVPEKNEGSMAKKFDR MTLMRAQLRARLASSKTSVPKQKEGLMAQKFDR MTLMRAQLRARLASGKMSFPKAPGQSKASPKAP RQRDEVMQDNMVEQDQGVHHTEDRVFSQVSG SPVLPGYKRMRTEEDESSSGPTTAFFLSDSRER QRRVVRSMNVLCRDADEDYLDSMDPTQEVNEM HQAWAEFSIRATSMMRRYRYFRELQDKVDELEA SAKKHVPEVATVNAKITRLEGMYNLADSQAKADV KRIASLKNRNNGKKKIAALAIKVAKLKEELERAKE SASASRQTHYDEWQKSEDGVTYLADVAQRSMDI GEKTTHDRMREALKRCFPEVDYNIVWDECEAIRL AEVDALAQEIYDLEHFRLY
808	sp_106560_tkfx	#N/A	0.876654	#N/A	autophagy 9 (APG9); CONTAINS InterPro DOMAIN/s: Autophagy-related protein 9 (InterPro:IPR007241); Has 489 Blast hits to 477 proteins in 190 species: Archae - 0; Bacteria - 0; Metazoa - 214; Fungi - 166; Plants - 51; Viruses - 0; Other Eukaryotes - 58 (source: NCBI BLink).	MSGWQGGRNVLSIFKWKKPGDPSLSIGLLNDNP PEIELSGYGRIPSPGSESPSELLNGESLKVEPIAD LDLFFERLYSYYCDKGLWCIIVKWIVELLSLAFTIC FSGFFLLYVDWDGLFNAKCGIDAVESGTKPCDL AIEALHAHPLTPLTLPKAIILGYLGIFSIYWIFCFLRF FAQLKEILRIRHFYYNCLHVTDDELKTMPWPSILE KVVRVQNSQKLCVVKDLSAHEIVMRLMRKENYLI GMINKGILAFPISKWIPGAGPTVKSGPNRGRQRLI LTKTLEWTLNWSILQSMFDRKFCIRGEFVSNPAT LRKRLVVVGLAMLLLSPFLVIFMLVYLFLRHAEQF YNHPTTASSRRWSNLSKWMFREFNEVEHFFRH RINSSIVHASDYLRQFPSPIISVVAKFISFISGGFAA VLIIIAFLDESLLEGHIFGRNLFWYAAVFGTITAISR AAVTDELLVLDPDGTMSLVVQHTHHMPKRWRG KENTEFVRVEFETLFQYTAMMWLEEIASIFLTPYL LIFVVPKQVNDILQFITEFTVYIDGVGDVCSFSVFD FQKHGNGNYASPYNSLPTRRSSQGKMEKSFLSF QTSYPSWQPDAQGMQFLTTLRTFRNEKLRGPIP GYAYSPPRRWQQQSPRFRDHSSRNSYFRRESP YNGQAPGCHLGSLWLIGTDPKTHPYLLDWFYVS GARHVIGNNTNDGPSVSLDIDNEQTNDFWMPSL NNEDNNNQGEESWPNFLDNRSKNHPEASTSSS PVFLEEESVLRHNDSSSTMLPSARSHWWNRRG GGTTSADPQQAASFLEPPEFHHRGGPHQASFLE PPMFRHSGGTQQASFLEPTEFQHRGGTQQTSFL EPTEFQHRGGRGPLQTSFLEPTEFQHRGGTQQT SFLEPTEFQHRGGPQQTSFLEPPMFRHSPEFNH HNADNYCESHSDRSIDDEHDEQPLDWGNYNQL SRTTDPDEFGGEFKLHFDDVYKPSEPPNMNLDH DKYHVS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
809	sp_106720_dzxp	#N/A	-0.8744456	-0.9163383	Protein kinase family protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase family protein (TAIR:AT3G13670.1); Has 17667 Blast hits to 17599 proteins in 1508 species: Archae - 20; Bacteria - 6581; Metazoa - 4779; Fungi - 1455; Plants - 2098; Viruses - 217; Other Eukaryotes - 2517 (source: NCBI BLink).	MPELRSGARRGRASTNPRTNNNQEEIGGQKAT VGRTRGRRGGAVAGGRGNKGGSRGRGRRNNQ QPAAAVGGGGDGCEGEADEKVVGQVKESSRKT ASLKAKEAAAEEVGVRKEIGEEEMDEYDSGGNA AGRSGDKGLNDDDGSTPLPEKVQVGGSPAYKV DRKLGKGGFGQVYVGRRSGGNTSERNGPGAYE VALKFEHRSSKGCNYGPPYEWQVYNALGGSHG VPRVHYKGRQGDYYVMVMDLLGPSLWDVWNNN SHSMSIEMVACIAIEAISILEKMHSRGYVHGDVKP ENFLLGPPGTPEEKKLFLVDLGLATKWRDSTTGL HVEYDQRPDVFRGTVRYASVHAHLGRTGSRRD DLESLAYTLVFLLRGRLPWQGYQGENKGFLVCK KKMATSPETLCCFCPQPFRHFVDYVVNLKFDEE PNYAKYVSLFDGIVGPNPDIRPINTDGAQKLIFQV GHKRGRLTMEEDDDEQPKKKIRMGMPATQWISV YNARRPMKQSFSPPPXVADIRLSQHIDKGNEDGL FISSVASCSSLWALIMDAGTGFSAQVYELSPHFL HKEWIMEQWEKNFYISSIAGAANGSSLVVMSKG TQYLQQSYKVSDAFPFKWINKKWREGFVVTSMA TAGSRWAVVMSRGAGYSDQVVELDFLYPSEGV HRRWDGGYRITSTAATFDQSALVLSIPRRKPNDE TQETLRTSAFPSTHVKEKWAKNLYIASICYGRTV S
810	sp_107050_dtss	#N/A	-0.8462243	#N/A	SAND family protein; LOCATED IN: chloroplast; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Vacuolar fusion protein MON1 (InterPro:IPR004353); Has 753 Blast hits to 594 proteins in 226 species: Archae - 6; Bacteria - 38; Metazoa - 289; Fungi - 213; Plants - 47; Viruses - 3; Other Eukaryotes - 157 (source: NCBI BLink).	MASDSTSSSSDSAYDSAADHNPNSDTPINNHDS AIDDHLTSLTLTPLDEDLAPEKEREIEETDQTPMP EIGVTEVEEEAADVLASEVVVEEERESNSVSATG DVGLAWREEEVEAPASPGSSGYAGERGSSNAS SASGIEEIGEDQVAERDGGLVNGVADSQPAWIP GKRHVNEDDTSILWRRRKKHFFILSHSGKPIYSR YGDEHKLAGFSATLQAIISFVENGGDRVKLVRAG KHQVVFLVKGPIYLVCISCTEEPYESLRGQLELLY GQMILLTKSVNRCFEKNAKFDMTPLLGGTDAVF SSLIHSFSWNPATFLHAYTCLPLAYPTRQAAGAIL QDVADSGVLFAILMCKHKVVSLVGAEKASLHPDD MLLLSNFVMSSESFRTSESFSPICLPKYNSMAFL YAYVQYLDVNTYLMLLTTRSDAFYHLKECRASIE AVFLKSNVLSEVQRSMLDGGMHIEDLPVDPSSR SGSASSHLGQHSPERFRDTYTSLGGPGCGLWHFI YRSIYLDQYVSSEFSAPINTPQQQKRLYRAYQKIY ASMHDKGFGPNKTQFFRDDNYVLLCWVTQDFE LYAAFDPLADKAMAIKTCNRVCQWVKDVENEIFL LGASPFSW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
811	sp_107230_jqya	-0.851934	#N/A	#N/A	Acid phosphatase/vanadium-dependent haloperoxidase-related protein; CONTAINS InterPro DOMAIN/s: Acid phosphatase/vanadium-dependent haloperoxidase related (InterPro:IPR003832); BEST Arabidopsis thaliana protein match is: Acid phosphatase/vanadium-dependent haloperoxidase-related protein (TAIR:AT1G67600.1); Has 1115 Blast hits to 1115 proteins in 408 species: Archae - 0; Bacteria - 729; Metazoa - 0; Fungi - 0; Plants - 228; Viruses - 0; Other Eukaryotes - 158 (source: NCBI BLink).	MTGRNIDVLLASSNSSTYSDQLLMFPPNFPLVSS LLAISLAQFLKIFTSWFKEKKWDSKRMLSSGGMP SSHSAGVSALAVAIGVQEGTSGSAFALAVVLACV VMYDASGVRLHAGHQAQLLNQIVCELPPEHPLS NCKPLRDSLGHTPLQVAAGAILGCIVAYLMQFYR
812	sp_107870_nafr	#N/A	-0.934834	#N/A	uncharacterized protein	MGRKITRSRKRGKVILNSEEEGNSVVDSSSRSE QDDSDDEDFKADGYQDLDDDDDEVVQLDVGSG VKRKRGRSKILYLDRLDKIPVQWGIFPRLKAWDM KEILEAKRAYRESGKKGDYGVLGVVIPVMDNEILR LKQRKKGNSERIPMKMRQMAKERAKKNSVKLER FELDMDTVNEEAIITPPPPRPESLSERCQQVTSS PPPQHKVTPPLPSPKVPPPQPKVTSPLRPPSLPK VTSPLPPPSPPKVTSPLPPPSPPKVTSPLPPPSP PKVTSPLPQSQANSTSPLPPPPLPKVNSSLPPPS QPKATSPLPHPSPPKVTSPLPHPSPPKVTSPLPP PSPPQFTLPRSPQLCVDPIGGANLEAVVNELMES PLPGSLDREKGKTIVCGLSAVPLGHDIDCDNKMD GLLASSALLFQRSRPPRRGRKKRVALADRGRGWF TSEIKKFSYNLVEFVNHNGGLLSGRGNLKGLRYV NYVVEKMKMCPVS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
813	sp_108470_hhjh	-0.9369839	-0.8738892	#N/A	Transducin/WD40 repeat-like superfamily protein; FUNCTIONS IN: nucleotide binding; INVOLVED IN: biological_process unknown; LOCATED IN: CUL4 RING ubiquitin ligase complex; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), BING4, C-terminal (InterPro:IPR012952), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat (InterPro:IPR001680), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: MOS4-associated complex 3B (TAIR:AT2G33340.2); Has 9286 Blast hits to 5748 proteins in 422 species: Archae - 16; Bacteria - 3151; Metazoa - 2206; Fungi - 2085; Plants - 594; Viruses - 0; Other Eukaryotes - 1234 (source: NCBI BLink).	MVVRAEKIVPAAEQDEEEVDVNVKKYLRGESAN LEVLQDKKLKGELAVKEELYKRSAKAAANVEKWL LPSEGGFLESEGIEKTWRIKQDAIGKEVDILSSKN CYDMVLPDLGPYKLDFTASGRYMAIAGRKGHLG VLDVKDMSLVKEFQVRETVRDVVFLHNELFFAAA QKKYPYIYNTDGVELHCLKEHGAVSKLQFLRNHF LLASINKFGQLRYQDVTMGKMIANYKSGLGRTNV MQVNPYNGIVGLGHANGSTTMWKPTSPNPVVK MLCHHGPLSALAFHSNGHLMATAGMERKIKLWD LRTFKVLQTMPGHAESLEFSQKGLLAASTGSFV QVFRDGSSSQNYSRYMSHSMAKGYKIEKVSFRP YEDVLGIGHSMGWSSILIPGSGEPNFDSWVANP FETTKQRREKEVRSLLEKLPPESIMMDPKKIGTLK PPRKKEKPTRKERFEEEMEAAVEAAKGKKLKNKT KGRNKDGKRAKKKQEVVQNAKKQFIEQNIKENE ELSKKKKRKRTSEEVAELPKSLQRFAKKKAD	MMGKEIQKETVNQGVCSGSSSPTKCDSSRASTA SQLSIDRDLGLPSCRVCQCVESDRRGDAALGFL GIDPLLVESHNVKELPEPGTKKALNDVEGSASVD TSKKLEPGYVQFISPTGEVFVCNTDLEMGACYNR DTLIELGCACKNDLALVHYACALKWFINHGSTVC EICGCIASNIRIADFNKVLASLKDYEALRDRTANG VPNFAHVHCNSDVSCVDPDAVAAIRRQRLSEIYL WFNPYGHLRHNNNTSENLNAPSVLVVTEPPANIV FQDVTPQNAATKWAVEGTGILLATGLLTVTLAWL IAPRVGKRTAKNGLHLLLGGVCALAVVVFFRCIVL TRIKYGPARYWAILFVFWFLVFGIWASRTHAAHT T

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
815	sp_109630_kwcn	-0.859905	-0.917186	-0.947545	polyprenyltransferase 1 (PPT1); FUNCTIONS IN: 4-hydroxybenzoate nonaprenyltransferase activity; INVOLVED IN: ubiquinone biosynthetic process, embryo development; LOCATED IN: integral to membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: 4-hydroxybenzoate polyprenyl transferase (InterPro:IPR006370), UbiA prenyltransferase (InterPro:IPR000537); Has 8781 Blast hits to 8779 proteins in 1934 species: Archae - 174; Bacteria - 4966; Metazoa - 236; Fungi - 294; Plants - 110; Viruses - 0; Other Eukaryotes - 3001 (source: NCBI BLink).	RKTANPPDSDQSFTITPSMAALCRLSRSTRKFLP FSLSYHSPHLSPSSTHPNSNPNYFHTRSPNHYTT PNFNFISNFWDHFTLCQKTFFSAYSTVPQGSKLA NSGKEDDDDEQLKSKQRGEASWVDLYLPRSAR PYAKLARLDKPIGTWLLAWPCMWSITLAAEPGNL PDFKMLALFGCGAFLLRGAGCTVNDLLDQDIDVK VERTKLRPIASGLLTPFQGISFLGVQLLLGLGILLQ LNDYSRILGASSLLLVFSYPLMKRLTFWPQAYLG LTFNWGALLGWSAVRGSLDPYIVVPLYVSGVCW TLVYDTIYAHQDKEDDLKVGVKSTALRFGDLTKE WLTGFGFVSISGLALSGYNAHLGWPYYACLTAA SGQMAWQIYAADLSNRADCNRKFVSNKWFGALI FCGILFGKLAA
816	sp_110690_skox	#N/A	-0.849669	#N/A	sumo conjugation enzyme 1 (SCE1); CONTAINS InterPro DOMAIN/s: Ubiquitin-conjugating enzyme/RWD-like (InterPro:IPR016135), Ubiquitin-conjugating enzyme, E2 (InterPro:IPR000608); BEST Arabidopsis thaliana protein match is: ubiquiting-conjugating enzyme 2 (TAIR:AT2G02760.1); Has 9584 Blast hits to 9566 proteins in 388 species: Archae - 0; Bacteria - 0; Metazoa - 4206; Fungi - 2037; Plants - 1834; Viruses - 20; Other Eukaryotes - 1487 (source: NCBI BLink).	MSGGIARGRLAEERKQWRKNHPHGFVAKPDTQ ADGSMNLLVWQCIIPGKAGTDWESGYYPLTMYF SEEYPSKPPKCKFPPGFFHPNVYPSGTVCLSILN EDCDWRPAITVKQVLVGIQDLLDQPNAADPAQT EGYHLFIQDTTEYKRRVRQQAKQYPASI
817	sp_110830_qnkm	-0.850701	-0.931246	-0.885115	AWPM-19-like family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; CONTAINS InterPro DOMAIN/s: AWPM-19-like (InterPro:IPR008390); BEST Arabidopsis thaliana protein match is: AWPM-19-like family protein (TAIR:AT5G46530.1); Has 177 Blast hits to 177 proteins in 19 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 177; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGTVGRNIAGPLLFLNLIMYLIVIGFASWCLNNYIN FSTHHPGLGGNGATSFFLIVSILAGVIGVVSKLAG ANHIRAWRNDSLAAAGSSALVAWALTALAMGLA CKEINIGGWRGWRLRVLEAFIIILTFFELFYVLAIH AGLFSSRYGPGYREHDYGMGVSGTTVDPKGTT TRV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
818	sp_110880_aptz	#N/A	-0.986387	#N/A	Protein kinase superfamily protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT4G10730.1); Has 125358 Blast hits to 123605 proteins in 4454 species: Archae - 114; Bacteria - 14309; Metazoa - 46798; Fungi - 12057; Plants - 31507; Viruses - 632; Other Eukaryotes - 19941 (source: NCBI BLink).	MGRSGHAIRSYTVNSNDYKLLEEIGFGASATVFR AIYIPYNEVVAIKCLDLDRCNSNLDDIRREAQMMS LIDHPNVVRAFCSFVVDRNLWVVMPFMAEGSCL HLMKTAYPDGFEEPVICSILKESLKALEYLHQHGH IHRDVKAGNILLDTNGLVKLADFGVSAQMFDSGD RQRSRNTFVGTPCWMAPEVLQPGSGYNFKADI WSFGITALELAHGHAPFSKYPPMKVLLMTIQNAP PGLDYDRDKKFSKSFKEMVAMCLVKDQAKRPTA EKLLKHSFFKHAKPPEFSVKKLLADLPPLWDRVK ALQLRDAAQLALKRMPSAEQEAISQNEYKRGVS AWNFDIEDLKAQAALVQEEEDLAEIKEADESVKS ATSSKDAPYYGSVSGNSNPPNNFINRKQGNVDE VAVTPTEGASCKIRTGNEFDFCSHGKNIVEKDGV DLGSGAPACEMDAAQVKVNPQVVKGRQSQSGP LMPGIVLGQTSTERSRVSERPENENQSAAEKSN RGSRRVPSFSGPLMLPTRASANSLSAPIKSSGGY RDSLDDKSKANLVQIKGRFSVTSESLDVVKDMM TSTSTSRRSSQGSPLRKSASVGDWISESKKTITS QPSGDINSGNVSASLLLPHLQNLFQQTAIQQDLIT SLLNSLQSTEGDGSQNGKPPAVPTSGSPENGIV ETVASERERLLLNKISDLQARMASLSEELSAEKH KHLLLQQQLNSVSSQEEDDEVVN	
819	sp_111520_konm	-0.980852	-0.982168	-0.95806	RESISTANCE TO P. SYRINGAE PV MACULICOLA 1 (RPM1); FUNCTIONS IN: protein binding, nucleotide binding; INVOLVED IN: defense response, plant-type hypersensitive response; LOCATED IN: plasma membrane; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: NB-ARC (InterPro:IPR002182), Leucine-rich repeat (InterPro:IPR001611), Disease resistance protein (InterPro:IPR000767); BEST Arabidopsis thaliana protein match is: Disease resistance protein (CC-NBS-LRR class) family (TAIR:AT5G35450.1); Has 20241 Blast hits to 16785 proteins in 626 species: Archae - 16; Bacteria - 1308; Metazoa - 4084; Fungi - 259; Plants - 14046; Viruses - 4; Other Eukaryotes - 524 (source: NCBI BLink).	MIISKVKELGFCQIPSNSEITEEFVQPHARRLSIHM DHSFHDADHQLMQSITKRKSSIRSLLYAKMYLQP QVSTNLFSKNFLLLKVLDFCNALIDYLPKEVGELL NLHYLSLRNTQIKRIPNSIGKLQYLQTLDLKGTPV YELPVDLNTLHKLRHLLTYCYKYDSGFSLHTQKL TGVKLQEKTLEKFGELQKLAFVDVGHTQTNSVV KELRNLRQLRRLGITELKFEDGKDLCTAIEEMKFL ETFSVYTNEFIDLDHFQSASELILKRLYLNGPLKQ HNTGWILELHSLVKIRLRWSRLEDDPMHLETLPN LVELQMLEAYDGDNLHFGNTGFKKLKILHLDLG NLKSLSIWKGALPLLQLMAIGQSMKLQVPLGIKHL SRLETLNFFDMPLDFARSLQHNGKFHSIVKHVPN VLFQKSDQGIWETCFL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
820	sp_111640_tmsu	-0.3201185	-0.871331	-0.9463832	CLIP-associated protein (CLASP); FUNCTIONS IN: binding; INVOLVED IN: mitosis, protein stabilization, cell growth; LOCATED IN: spindle microtubule, plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: HEAT (InterPro:IPR000357), Armadillo-like helical (InterPro:IPR011989), HEAT, type 2 (InterPro:IPR021133), Armadillo-type fold (InterPro:IPR016024); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MEEALELARAKDTKERMAGVEHLHQLLEASRKP LTSSEVTSLVDTCIDLLKDNNFRVSQGGLLALASA AVLSGDHFKIHFNALVPATVERLGDAKQPVRDAA RRLLVTLMEVSSPTIIVERAGSYAWEHKSWRVRE EFARTVTSAINLFASTELPLQRAILPPILHMLNDPN HSVREAAISCIEEMYMQISFKMSYKDTSFHHIW
821	sp_112110_iddz	#N/A	-0.8713325	#N/A	mediator subunit 8 (MED8); Has 21670 Blast hits to 11104 proteins in 489 species: Archae - 4; Bacteria - 472; Metazoa - 8306; Fungi - 2813; Plants - 2364; Viruses - 51; Other Eukaryotes - 7660 (source: NCBI BLink).	MEAMMGTQQGQTQAPPQQPPQPEKVVEKLNSA VQQQLNLDSVKTRAISLFKAISRILEDFDVLSRSN ALPKWQDVVGQFSMVNLELFNIVEDIKKVSKAFV VHPRNVNAENATILPVMLSSKLLPEMEVDDNSKR DQLLLGMQNLPIPTQIDKLKARIDMIAAACESAEK VIGETRKAYGLDTRQRMSTVPTLDKAFSAKIQEK ENLLRAAVNHGEGLRLPVDQRHVTSPLPSHLAD VLTVGDGGQTYGDGSGGMYSKNTPVISGNHNNQ GLPLQMQASGGQLLGRSVPSPSTATVASTAENT ASPMQYANSPRSGTNMINAPSSQQQTLQQQHQ LQQQRKMMQLPTQQQQQQLLAQQQQFRQSAMQ GLAQLQGQSQMQFSQPLASQQFQSRQLPTGHI GQSQLSQGNQLNRHIGQFSGAANTALFNASQNS PNASVMSMPATMTSQSMLPRMQQFALPGGNPQ RTHPSQMLSDQMFNMGANQASMMSLPQQQQV PQGAFGSMSQNSQNLQSNMVPLQNATQNLQNF CQQQRQQNQQ
822	sp_113170_cfim	-0.3463383	-0.9453397	-0.9760552	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion, endoplasmic reticulum; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF106, transmembrane (InterPro:IPR002809). Uncharacterised conserved protein UCP010045, transmembrane eukaryotic (InterPro:IPR008568); Has 409 Blast hits to 409 proteins in 203 species: Archae - 0; Bacteria - 0; Metazoa - 152; Fungi - 136; Plants - 53; Viruses - 0; Other Eukaryotes - 68 (source: NCBI BLink).	MAEDLVLDTTIRDWVLIPLSVVMVLIGILRYFVSKL IRTFQNPDGKIVKEGQVVIRARNLRSSADFIPPKS FRARRTYYSNEENGILHVPKGQAPNPQAQIMFSD PNMAMDMMKKNLSMIIPQTLTFAFVNFFFSGFVA AKIPFPLTQRFRSMLQNGIDLSTVDVSYVSSRSW YFLNLFGLRGLFSLILGEENAMDDTQRMWQMGG FGFDPSKSLGAEKDSLDIVQHDWVLPKFEQRAE AVLRKLAK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MKRTTPTAGEQSKGQEAQMDPAESTGFDPTVDIS
						SQLMQRYSSSSAPQHRHLMATAAAVKSILLEDSL
						PLSPLCYFAATINALSQPQNDAASASALATFLSIV
						LSLVPEDYINPEKAAEAAAILVRVLGSEEGEMVV
						GSASSSARCVVKCLGILLGFCDLEDWDSVELPFQT
						LLKWSIDKRPKVRKCALVYLENALKSFQSETVKR
						KASILVMSLLESYMSIAVKLGIIQKVDESGSERQC
						ETEQLDVVHALSVVRVALPYICDKSIKKTILKLIELL
						DSRSSTFTRPVLNAIEALLEKVEGGIIVQVAEDIID
						ALSSFISSETNPNDSIISAANLLKISLDKLHSEDTD
						KWNKSLSVGVNAIAGLLTREDEVASKSSNILKNLI
						VDQLSDDIESQPANGEESKGITSMCSTFMEMLD
						SLAGIPNEHTLAVLSVLFLKLGRSSCSHMKNIVLK
						LAEIFTHADGNKHDTVHLQNSLGCAVVAMGAEN
						VLRLVPISFDEENSTYANMWMLPILKKYTCGASL
						GYFTEHIVPIAQSLQKASRKVKKSVTGQDLQALA
						HGLWGLLPAFCRHPTDVSKSFQSFAKLMLVQLK
						KDASMREEIANSLQELVKQNKCLPKFDAAIVEPIK
						HSSGFSVDDYLVEKRNLLSYSKKAAIKNIKALSLY
					ARM repeat superfamily protein; FUNCTIONS IN: binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-type fold (InterPro:IPR016024), Domain of unknown function, NUC173 (InterPro:IPR012978); BEST Arabidopsis thaliana protein match is: ARM repeat superfamily protein (TAIR:AT2G34357.1); Has 378 Blast hits to 366 proteins in 171 species: Archae - 0; Bacteria - 6; Metazoa - 125; Fungi - 137; Plants - 63; Viruses - 0; Other Eukaryotes - 47 (source: NCBI BLink).	SNELLLALMTVFLNSKPAKRSSLKKAISCLASISD
						SSVTKKIFISSLERFPSVKAVSVSEDSDPSNALLS
						DVPSDSSTIGENSQWSLMLDLASSIVERADEDLV
						GLIFKLTKHVLQAGDGKCITEAYYTLSCVLEEHP
						WFCSLKMDELVDLIMGLKAPADISSFSNRLSCLH
						LLLVHSLKSSSDVENTKPFLILNDIILALKDSREEA
						RKIAYDVLLKLSSCLVNSSDADSDGPYYKLISMIM
						GYLSGPSPHITSAAISALSMFVFKEPELCVKVPHV
						VSSVASLLQTKAVETIKAVLGFIKVLVSSLEAKDL
						QNFLPAIIDGVLPWSSISRHHFRSKVTVILEIVIRK
						CGFPTIKSATPERYQSFIRKVSQNRHGNTTSKEA
						DSADAKTQLSNVPINGSKKRSYEETNNNRNGSE
823	sp_113760_euny	-0.8811083	-0.8226046	-0.9868835		LQRNKRRNQIQDFQSPGGTRRNVAGFDKKPNKF
						MEHRWDFLDSLDFDVSSKILMGLDDFADIVPASS
						VSRLWRDFVITNGICKHLCLRLFPQLSGIDRVDI
						SNKQENSDVGSSSFVELEHLKKEHRAFVALARG
						LTSSEVGECLSDAISATSTDNYPEEGVHNTLVSR
						DRIGRRASYWSSSGQSNPEVPERLTYKLISDLCII
					F-box family protein; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364); Has 80 Blast hits to 80 proteins in 29 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 11; Plants - 57; Viruses - 0; Other Eukaryotes - 12 (source: NCBI BLink).	TEMNVHPFQAFFQSNSPIYSAKAVRFRMGHRILP
						EGDEEDIVHEGQECAEEKFVWTYTSPEFPMAQE
						NRVQNFKLPEPILCFGGYLQIELLGRVQRQEMDG
						LFYICVAHVQVMGRSLFPVFGVESLDPNGRFTLT
						YNPHAQFSSPHNGNEESDEISVTPEMVQRHVRG
824	sp_114360_gqqp	#N/A	-0.8822397	#N/A		WEQILNMLRGTLGVEVYDSDDEQHDSEDEMPEE
						FAL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
825	sp_114370_keez	-0.8857865	-0.8673487	#N/A	phloem protein 2-A13 (PP2-A13); FUNCTIONS IN: carbohydrate binding; INVOLVED IN: N-terminal protein myristoylation, response to wounding; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364); BEST Arabidopsis thaliana protein match is: phloem protein 2-A12 (TAIR:AT1G12710.1); Has 444 Blast hits to 438 proteins in 25 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 444; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MGVGMSSTDGGNHGGETPEYRRKLEDIPESCIA MVLGHLDPPEICRLARLNRAFRGASSADFIWETK LPTNYSFIVNKISELNHKNDDDMKNYGGSKDIST GLNKKDIFARLSRHNLFDSGNKEIWLDKKTGGVC VSISSKALSITGIDDRRYWNHIPTEESRFQTVAYL LQIWWFEVCGELEFKFPAGTYSLFFRIQLGKSSK RFGRRSCNVEHIHGWNLRPVRFQLTTSDGQHAS SECYLDSPGSWILYHGGDFVVKETDKLTKIKFSA RQIDCTHIKGGLSFDSVMILPKSIASKAR
826	sp_114620_aksi	#N/A	-0.8730363	#N/A	VQ motif-containing protein; CONTAINS InterPro DOMAIN/s: VQ (InterPro:IPR008889); BEST Arabidopsis thaliana protein match is: nucleotide binding;nucleic acid binding (TAIR:AT1G21320.1); Has 294 Blast hits to 294 proteins in 63 species: Archae - 0; Bacteria - 0; Metazoa - 108; Fungi - 21; Plants - 144; Viruses - 0; Other Eukaryotes - 21 (source: NCBI BLink).	MDHPDSPGGGRPPRKELHGPRPAPLKVRKESH KIRKPPVAPPAYPTQNQQLPPYLPPPRQPVIYTV SPKVIHTHPSEFMTLVQRLTGSDATSSSYATSSN VTSSSSTIVDGFSSSTTTTRSHHVTGGISPAARFA SIEKTKMHSEAKKFVGPSEVFRREVMDIEEYVNS GGGGGGGGGGGGSGIEGVERTMVFPVEIACVEA
827	sp_115131z4387	#N/A	-0.8725285	#N/A	uncharacterized protein	#N/A
828	sp_116390_zmwk	-0.8936887	#N/A	-0.8593714	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G11290.1); Has 881 Blast hits to 512 proteins in 30 species: Archae - 0; Bacteria - 2; Metazoa - 0; Fungi - 38; Plants - 833; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).	MNEQKVKRSKATWDDESTRIFCEVCADEVHAGN RPNTHFNRVGWDNVVKKFQLRTGKRYDQKQLK NKWEKLKSEFSTWKNLHKETG-GWNHDNNTID ASEDWWENKEQVDANAAIFREGGIKNLTEQEIM FSKIYASDATSWNPYSEEEVVEETEVEVEDDSIV GNFDVEDEEQVDITLDSNTTQTHPLQPESENIRK SNVTIPTGVKVKNVCKGKKSKVSTAKLMQNELKR IVGAMESFSQSSATVTSRRTDLPGCSIQECLTLL GVTPGVEVGSELYMLGTRLFMKKECREMFVALP GEDIRSAWLDQELLREKEKKNKT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
829	sp_116910_xyir	#N/A	-0.6739836	#N/A	Oxoglutarate/iron-dependent oxygenase; FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, iron ion binding, L-ascorbic acid binding; INVOLVED IN: oxidation reduction, peptidyl-proline hydroxylation to 4-hydroxy-L-proline; LOCATED IN: endomembrane system; EXPRESSED IN: petal, sepal, male gametophyte, flower, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Prolyl 4-hydroxylase, alpha subunit (InterPro:IPR006620), Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123), Metridin-like ShK toxin (InterPro:IPR003582); BEST Arabidopsis thaliana protein match is: Oxoglutarate/iron-dependent oxygenase (TAIR:AT3G28480.1); Has 2477 Blast hits to 2439 proteins in 330 species: Archae - 0; Bacteria - 383; Metazoa - 1060; Fungi - 71; Plants - 412; Viruses - 15; Other Eukaryotes - 536 (source: NCBI BLink).	MKFDIICRSSLLHSSVSVLLLLVVVPSMLLLCSSW SEVLVRCSLLLVSFLFNLLVSFLVFVYRGFITEEEC DHLIPLAHGQKEGTNSVISDTQGVSARSLVTLDL
830	sp_117090_ztsk	-0.9853147	-0.9931851	-0.9903314	CONTAINS InterPro DOMAIN/s: P60-like (InterPro:IPR011687), Tumour suppressor protein Gltscr2 (InterPro:IPR011211); Has 709 Blast hits to 643 proteins in 201 species: Archae - 0; Bacteria - 32; Metazoa - 224; Fungi - 154; Plants - 45; Viruses - 0; Other Eukaryotes - 254 (source: NCBI BLink).	MGKNRAKKVKRAWRNISTKEFDDYNEKATRDAN SGGSLASAPSDSLFVVDKSTDVPVRRKIEKKREK VLRFESVLQRNPFVKAVPSSILKKLKKKDIKTVDLA QDASKDDDKKGPESGKFDIWDDQGKVNIKGKKK ANPSLIPAVEVEPPGCSFNPENESHKDSLALAVA EEMKKIYKKELEPEPIPLTVPGEPVTEDDRYFLEV DHEDDDVDEMDEENQTEDQINEDENRKSKEKR ITTVVKNKRIRRKEQQKAEAEAKKIKDLSKEIDSIP DILQEIAKEDEEKKNRHLRRMVAKEERLKSGPPR LGKRKFEPAPSQVLLTEEITGSLRKLKACCTLAR DRFKSLEKRGILVPSAKFSRK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
831	sp_117540_guqq	-0.385625	-0.987486	-0.977733	arogenate dehydratase 2 (ADT2); FUNCTIONS IN: arogenate dehydratase activity, prephenate dehydratase activity; INVOLVED IN: L-phenylalanine biosynthetic process, metabolic process; LOCATED IN: chloroplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Prephenate dehydratase (InterPro:IPR001086), Amino acid-binding ACT (InterPro:IPR002912), Prephenate dehydratase, conserved site (InterPro:IPR018528); BEST Arabidopsis thaliana protein match is: arogenate dehydratase 1 (TAIR:AT1G11790.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MLDVASSNSLTSSLIPQFRVQFTQPSHHYATLN FPKTHFKRCRNLIDSSLHPTSNYIPQAILSQKLPE DSPSAADATNPLPRPLTSTVASNSDGSSLRVAY CGVSGAYSELAAETAYPNCEAVPCDQFDTAFEA LERWLVDRAVLPIENSLGGSIHRNYDLLLRHRLHI VGEVKLPIRHCLLANPGVKIDDLKRVLSHPQALS QCEHTLAKFGVTREAVDDTAGAAKFVAFNKLQD AGAVASVAAARIYGLNILAEDIQDDSDNVTRFLML AREPIIPGSDRPFKTSIVFSHTEGPGALFKALAVF ALRQINLTKIESRPLRNSNNGSSKCFDYLFYLDFD ASMADPNAQNALRHLGEFATLLRVLGSYPADTS RR
832	sp_117640_yejk	#N/A	-0.910784	#N/A	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein; FUNCTIONS IN: methyltransferase activity, catalytic activity; INVOLVED IN: metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Methyltransferase type 11 (InterPro:IPR013216); BEST Arabidopsis thaliana protein match is: S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (TAIR:AT3G60910.1); Has 2390 Blast hits to 2390 proteins in 771 species: Archae - 92; Bacteria - 1226; Metazoa - 381; Fungi - 88; Plants - 206; Viruses - 0; Other Eukaryotes - 397 (source: NCBI BLink).	MVVPSMSEIKAMVADMLDLPFADDSFDVVIEKGT MLL
833	sp_117670_ndhh	-0.880367	-0.914029	-0.952377	Protein of unknown function (DUF803); INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF803 (InterPro:IPR008521); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF803) (TAIR:AT3G23870.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVMSSDNVKGLVLAVSSSIFIGSSFIIKKKGLMKA GTTGTRAGSGGHSYLYEPWWWAGMITMIVGEIA NFAAYAFAPAILVTPLGALSIIFSSVLAHFILKEKLHI FGMLGCVLCVVGSTTIVLHAPHERDIESVKQVWT LATEPGFLVYAAIVVVMVLVLIFYYVPRNGRTHMI VYIGICSLMGSLTVMCVKAVSIALKLTFEGNNQFK FFQAWFFTIVVIVCCILQINYLNKALDTFNTAVVSP VYYVMFTSLTILASMIMFKDWDSQNGSQIATELC GFVTILSGTFLLHKTKDMGSTPAAAAVPVTELVR QYSSNCIRDSSKHSDRHGSPDRNCRE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
834	sp_118770_kxzo	#N/A	#N/A	0.85421	U3 ribonucleoprotein (Utp) family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: rRNA processing; LOCATED IN: small-subunit processome; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Small-subunit processome, Utp14 (InterPro:IPR006709); BEST Arabidopsis thaliana protein match is: U3 ribonucleoprotein (Utp) family protein (TAIR:AT5G08600.2); Has 7468 Blast hits to 4514 proteins in 366 species: Archae - 14; Bacteria - 554; Metazoa - 2650; Fungi - 816; Plants - 392; Viruses - 146; Other Eukaryotes - 2896 (source: NCBI BLink).	MADNKRKSSGGAKQLQRREKTPKKKFSSKKSNPK SSGKRRGGGSGGGGERRGPRLPSAFRKQLDNF KNPNNGDDDVEDDEVIDSDEGELNYVKDFVVYE EEVAEEESKKNRRFDHVENYEYQLPEKFKDEDIL SDDDEDDDLDQDKEDEDSVDKEGDDGRHARML TDITGIHTEAFAGKKRKRNVVTEAYPESEFNPSS DVLSGNGKISIQDLLDPLRGTSSYSKLRKRIEHVE KKGETVQAPLSKPDRERLERGVAYDHVKKDITK WEHLVKRNREAPTLYFDQGNDMGLPTVGTMAS EFKPRSDFEKAIASIMDNKEVAEAYQSDGAKLLE LNKISVEEVKDRQDRLAKMRSLLFREHEMKAKQVK KIKSKTYHRLKKKDKLKAASSDPQTDPEAAREYA EKQERKRAEARMTLRHKNQGKWAKRVLERGLS KQDEGTQAAIAEQLHLHNLLTRKMNSVKENSSS DDSSCEDDEDMDSSDEEGASKLLAKAKEKTIKL MEQEDEVPDSGVLSLPFMVRGLEKRKEAANGEA KRALDDYELSLNQTAEKGLKNGTPSGRRVFGVT KKVEEANIKHRPDVDNYGSDNEDTIEEKDDIHME CAKGNHMEKDVDIDLDVLREDLETGDDSAFKSVI DVAEDSGPRTSRDMKNDGCKGGSTHDANVKSK PEVAKPSPSSDSGEGEDDSDTDSGGHMIDGMLS SGARQTYELPTNEELIKRAFAGDDVEEEFEKIKE DVLNEENPEPEKPLLLPGWGQWTDVQKKRGLP SWMLKDHANAKQKREEALKKRKDAFLKHVIISEK IDKKVSFLIFDKYHLPLLVLKFTWWVKSLDAGQV VN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
835	sp_119020_netp	-0.947759	-0.955262	-0.902571	INVOLVED IN: protein localization; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Vacuolar protein sorting-associated protein (InterPro:IPR009543); BEST Arabidopsis thaliana protein match is: calcium-dependent lipid-binding family protein (TAIR:AT1G48090.1); Has 2219 Blast hits to 1322 proteins in 202 species: Archae - 0; Bacteria - 2; Metazoa - 1077; Fungi - 361; Plants - 231; Viruses - 0; Other Eukaryotes - 548 (source: NCBI BLink).	MFEGPVRQLLLGYLGKYIKDFHQKQLKITLDEIVL KDVELILEAFDYLQLPVDLKRCHVGRLRIGMSWK KLSWESWDPIVISLELEDVLFCAGPRDDHEWSID AVERRELAAKKAKLAAAELEKLSRRVCDNQAGQ SLISHVTGRILDSIQVSMRNVHVVYCDIQSQSMLF GLRFSSLEMKQAADGPLSSKARGGQVNKIIEIRG LGIHCSTLEGFLGSLGVNGVGETELWRRVVTDV DKYDYLLMPFDLSLQVTKSGKLEFNAPQYSVVSE LTKLVMTINPAQLQHILMFSDYICTCRLREKYGRY RPWDFPLSKKLKGWQKSWWHYAKESVLYDVRK RRKETSWRHLGQKIVQLRKYVNLYKVKLDFLQR EQLVDQDTLLKLEQMEKDSDIEDILRFRSYAEREL EDSQRSVSSDAIEGSTPAENLPNDEVPTGGARG WLNWLSRGMLGAGGTDDSSQFSGVVSDEVLKD ICKATKFQPMPSVTVAANGVKFLFAVKFHVSKMS ILINSTKLDQEIAELSCFATQIEFNVWEESTKILAK VESAEVIDSSKQMAILQMQKVSTCQNVLESKKPF ASTKVDISVKNQMEEFLFKVMIQPVEIYCDLKFLL RFVEFYSVMESHKSLQERVLLSLNGIEDANCRLQ SKAE
836	sp_120190_wezk	-0.862549	#N/A	-0.857536	unknown protein; Has 157 Blast hits to 144 proteins in 62 species: Archae - 0; Bacteria - 0; Metazoa - 101; Fungi - 0; Plants - 35; Viruses - 0; Other Eukaryotes - 21 (source: NCBI BLink).	MAEEERELPLTMDKDTNLLSKVAANHLYLTQFEA FRSTILSLRHRNPNLALSLLQTIVVNGGNFKNAIY STNCSSPALLTWLSSLELFQFENCTSIWSSSTSA DSLRLRVEFLLYIQMVSSRVSESVNRKAEMEEEE AEEELRELKGSMLLLDGIMELGLRRLKEDLIVSLG EERGSVEEEGVVIREEELGCLKRVVLKNADVFVS LCENIQGQVGVELEKDDDNELAITLRREGRVGTV LSEEDQRVLQSIQKYAQIAHLEEMRQCVIMDDVD EAILHIRYLHLVYGVSKEEYRMVFQDLLQKVFSGK EGLGDTSCTTRDKLLYIYREALLSNCTHLVQMIQP CW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
837	sp_121190_ouox	#N/A	-0.85332	#N/A	PHD finger family protein; FUNCTIONS IN: zinc ion binding; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, PHD-type, conserved site (InterPro:IPR019786), Zinc finger, PHD-type (InterPro:IPR001965), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: homologue of trithorax (TAIR:AT2G31650.1).	MTDNSRCQGLKKMMGRGAEGGCGTEEKACPVT RSSPPASSPAPVSSPENLEKVNSGVDLYSQAFK VLSERSPFDSDEVLAKTVATLPSGLAGLLSKQSD SKKKHKRPHSETKSSRKGSTEKSKSSRLWTGLE DYFRELTLPDIERLYRYSPSFNSILNSSTSNSSLK VPVLGNGVRTSSDNDNVSKIENRHANAATAAAV KDEEQDEVLDANVPLLDTGALQDKEEFDKLHDV KVENGSRDSLSVEKKGPFPEYSLQSCSSMEWVL GSRNRVLLTTTRPTKKRKLLGVSAGLERLMVAQ PCEGSSSLCHVCSMGETGDLLNQLVVCNMCNV AVHQKCYGVQGEVDGSWLCSWCKHKVNARAG PSSGNDIPCVLCPKSGGALKPFQNARSGDGGSV EFSHLFCSMWMPEAYIDDTRLMEPVLCRGMKDV RGKLVCNLCKMKLGSCVRCSDGSCRTSFHPLCA REAKNKLEIWGRFGCDDIELRAFCAKHSGNLDLD DSVALESHEQNHTVGSDSLGSSIESALLPGNKIK KLKIGCKNGDKVTFQMRISDAKSNDLGVRERQDT GIKDADSNTKLKSGSGVEEYEDVSTQLSDIGLGG TGDDVKTSRVLDLGLVLKKLIHRGKVSLKQIAAEI GIQEASLASRLADNHLSPDLYCKLLKWLKGRSYL DTAEKDWRAKLSCLVPSKAELKTDDNTDAVAVR DLDISDVPVKSVPPRRRTVGNIRILKEDNAVSLCI DKSLGNGLITKEGTDPEDTIAEHDNLNKDSNLDG MEEIMVETNGFGHPTAEPSPTSDSTLLKPNSLFD AERNGAEEGTSSGKSSLLDLDQETPSCSSAYRV GDCDGNIEPGINSSVQPYNQKKLDEEQSLLSRRT PDFNGCIDREPLAMEQTLSVKKKFPACIDEVDGG EGMKLEDLRKADEVVVSDFSPENDVQREIIFFQQ QLSHNAVALKNHSDQLVCKIVEKLPNEMDSVRKR RWDSVHVNQYIYELKEARKQGRKERRHKEAQAV LAAATAAAAASSRVSSLRKDAVDESSQQENQLK VEIGRVRGSLSSLSVPRAKVSNPSLDVPKAISDG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
838	sp_121220_xmqe	-0.8362441	#N/A	-0.8545831	VPS9A; FUNCTIONS IN: Rho guanyl-nucleotide exchange factor activity; INVOLVED IN: cell plate assembly, transport, cell wall biogenesis, post-embryonic root development, embryo development; LOCATED IN: cytosol, nucleus, membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Vacuolar sorting protein 9, subgroup (InterPro:IPR013995), Vacuolar sorting protein 9 (InterPro:IPR003123); BEST Arabidopsis thaliana protein match is: Vacuolar sorting protein 9 (VPS9) domain (TAIR:AT5G09320.1); Has 1001 Blast hits to 992 proteins in 208 species: Archae - 0; Bacteria - 4; Metazoa - 518; Fungi - 219; Plants - 81; Viruses - 0; Other Eukaryotes - 179 (source: NCBI BLink).	MENSEMFSSSTVQHTWHDFLDRMRQPSAAEFV KAIKSFIVSFSNNPPDPEKDSAAVQKFFANMEVA FRAHPLWTGCSEEELESAGEGLEKYYMTKLFTR VFASFPDDVKTDEQLSEKIALVQQFIRPENLDIMP TFQNESSWLLAQKELQKINMYKAPREKLACILNC CKVITNLLHNAAVTANENTPGADEFLPVLIYVTIKA NPPQLHLNLQYIQRYRRESRLVSETAYYFTNILSA VSFISNIDAQSISMDENEFERNMESAQALLSGLTT DFQTTPNAVDNNTSSSKPPMEEFQQTLPSNEND STKQPNPPNPTQPEPNPMTKNMSISDLEEKGAT TILKEDRSINNVFQEPPYVFVSVGDLTLRDVGDLL NNYKQIVFKYVCLAKGLGVTPSLSTSSVSQVDQQ SLVEKTTEGSEPKRETETETDEQLCRKQDSSKT DTGISESKSPEKESDEVLVSLKEDDSEGNQVDIA ATTTKESQDDVQ
839	sp_121450_ixur	#N/A	#N/A	-0.8305525	Cytochrome b-c1 complex, subunit 8 protein; CONTAINS InterPro DOMAIN/s: Cytochrome b-c1 complex, subunit 8 (InterPro:IPR020101); BEST Arabidopsis thaliana protein match is: Cytochrome b-c1 complex, subunit 8 protein (TAIR:AT3G10860.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGKIPVKIMKMVQYALSPYQQKVMPGLWKDLPG KLSHKVTENYISTTLLLGPLVGVYSYVQWYLEKE KLSHRF
840	sp_121460_jegf	#N/A	-0.8693181	-0.8663862	RNA-binding (RRM/RBD/RNP motifs) family protein with retrovirus zinc finger-like domain; FUNCTIONS IN: RNA binding, nucleotide binding, zinc ion binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif., RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677), Zinc finger, CCHC-type (InterPro:IPR001878); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein with retrovirus zinc finger-like domain (TAIR:AT5G04280.1); Has 996832 Blast hits to 989721 proteins in 35749 species: Archae - 21265; Bacteria - 589617; Metazoa - 208621; Fungi - 26263; Plants - 58603; Viruses - 68007; Other Eukaryotes - 24456 (source: NCBI BLink).	MAARDGYRIFVGGLGWNTSQRHLEDAFVRYGKII DSLVMVDRETGRPRGFGFITFADRRAMEDAIRS MHGRELDGRVISVNKAEPRMGAEDPYDAYGGD RRSGGRESYRGGERPGERSDECFKCGRAGHW ARDCPSNGGGRYSSQSEFGRSGLQGDRDYDRH YDRAGGGRYDSDRDLESRDVYVSRDRYSTERY PPAGDRFGSDRFMDRYPQNGYVRDRDYERDGG IRGGDRYAAGGPTRYDKGSYRDRVHPYDAPRRT SRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
841	sp_121550_jwro	-0.961307	#N/A	-0.873958	casein kinase I-like 10 (ckl10); FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: casein kinase I (TAIR:AT4G14340.1); Has 56550 Blast hits to 56234 proteins in 2232 species: Archae - 39; Bacteria - 9117; Metazoa - 21673; Fungi - 6267; Plants - 8804; Viruses - 331; Other Eukaryotes - 10319 (source: NCBI BLink).	MDHVIGGKFKLGRKIGSGSFGELYLGINVQTSEE VAVKMESAKTKHPQLHYESKLYMLLQGGTGVPH LKWYGVEGEYNVMVIDLLGPSLEDLFNYCTRKFS LKTVLMLADQMINRVEYMHSRGFLHRDIKPDNFL MGLGRKANQVYVIDFGLAKKYRDLQTHKHIPYRE NKNLTGTARYASVNTHLGVEQSRRDDLESLGYV LMYFLRGSLPWQGLKAGTKKQKYDKISEKKMLT PIEVLCKNYPSEFTSYFHYCRSLRFDDKPDYSYL KRLFRDLFIREGYQFDYVFDWTILKYPQIGSSSRP RPTAKPALNFGPSAERAERTSGQDIFDRFSGAV EGFARRNNPGTGSQGDQSRHRSADDPSSKDVQ ADRDRGRISRNGSTSKRAIASSSRPSSSGEPSDS RSSRLMSTSGRLSTTQRVQPGFESKTSTLARVS AARPGRDDALRSFDFLSIGAGRRK
842	sp_121670_qnej	-0.956341	-0.313723	-0.938685	Transcription factor TFIIIC, tau55-related protein; CONTAINS InterPro DOMAIN/s: Transcription factor TFIIIC, tau55-related (InterPro:IPR019481); BEST Arabidopsis thaliana protein match is: Transcription factor TFIIIC, tau55-related protein (TAIR:AT1G80745.1); Has 31 Blast hits to 31 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 31; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MELDVTTIDHNQQDEYVLLNLDAVCGQVNILLNE PYVLTGLDTANPVLTIGDKLKLIGEYEETIGTCLIF SDKVLYEETGSSERNLFKGACIVDPDQGPSKQIK PVTSLHKILKFRLLLETDMPQKSPDLPRQEVN
843	sp_121690_zyjp	-0.932423	-0.958313	-0.987452	ankyrin repeat family protein; FUNCTIONS IN: methyltransferase activity; INVOLVED IN: metabolic process; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ankyrin repeat-containing domain (InterPro:IPR020683), Arginine N-methyltransferase 2 (InterPro:IPR017408), Methyltransferase type 11 (InterPro:IPR013216), Ankyrin repeat (InterPro:IPR002110); Has 14669 Blast hits to 7820 proteins in 738 species: Archae - 65; Bacteria - 1760; Metazoa - 7798; Fungi - 1170; Plants - 439; Viruses - 38; Other Eukaryotes - 3399 (source: NCBI BLink).	MEEKGEKLCEAARIGDVAGVTSLIDAGYDVTYFD ADGFTPLMYAAKLGHHEVISKLLEAGAPWNALSP SSLSAGDFAMQNGHQDAFDLLLNTGVQAELILGT IARKEKENSHSDGGYLVDRVSFSEDKLMDADSK AVMMAWEKPLMESHAKAVCSGGGHVLNVGFG MGLVDMAIQQYNPVSHTIIEAHPEVYDRMLRNG WGEKENVKILFGRWQDVLPQLGSYDGIFFDTYG EYYEDMREFHQHLPKLLKPGGIYSFFNGLCGDNA FFHVVYCNIVSLELENLGYTTQLIPLPVKDCLAEE VWEGVKQKYWQLDTYYLPVCQYTDDSE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
844	sp_121800_mdna	-0.923832	#N/A	#N/A	alpha/beta-Hydrolases superfamily protein; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF676, hydrolase-like (InterPro:IPR007751); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT1G10040.1).	MLLQTRVDTPFTIVFLSAADWKFGVEQFVKMLPD KVHVHRGKETGVLVREGTRPTSVRAMWPLEFPS
845	sp_121920_kswo	#N/A	-0.951311	#N/A	uncharacterized protein	MTEFFPDEQLLSLCRIDTPWYADMVNYLVTRTIP SYIV
846	sp_122360_dlqq	#N/A	-0.898374	-0.862443	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF2365 (InterPro:IPR019314).	MEEGGKEETLSIENSSMKEKETMEASKNFNQQD LFKALEVVERDSLAIAQSYTALFASLRSALSEVTS TSVDHIQCFSDAAGKLQES
847	sp_122450_neft	-0.922304	-0.955709	-0.922373	LATERAL ROOT STIMULATOR 1 (LRS1); FUNCTIONS IN: nucleotide binding; INVOLVED IN: lateral root formation; LOCATED IN: CUL4 RING ubiquitin ligase complex; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), Protein of unknown function DUF3337 (InterPro:IPR021772), WD40 repeat, conserved site (InterPro:IPR019775), WD40 repeat (InterPro:IPR001680), G-protein beta WD-40 repeat, region (InterPro:IPR020472), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: Coatomer, alpha subunit (TAIR:AT1G62020.1); Has 59803 Blast hits to 26869 proteins in 729 species: Archae - 50; Bacteria - 7042; Metazoa - 23770; Fungi - 14096; Plants - 7202; Viruses - 3; Other Eukaryotes - 7640 (source: NCBI BLink).	MHRVGSAGNTSGSNRPRKEKRLTYVLNASDDTK HSAGVNCLAVLKSNDMCNSEYLFTGSRDGILKR WAIEEEAATCSATFESHVDWVNDAVLVGDHTLV SCSSDTSLKTWNCMSNGVCTKTLRHHSDYVTCL AAAGKNSNIVASEAALAPVSKSNDSTEEEASNGI NGGGSNLPITSLRTINSSNSLHSTPSHGYVPTPA KGHKESVYALAMNDSGTILVSGGTEKVVRVWDP RSGSKTTKLRGHTDNVRALLLDSTGRYCLSGSS DSMIRLWDLGQQRCVHSYAVHTDSVWALASTTS FSHVYSGGRDQSLYLTDLSTRESILLCTEDYPILQ LALDDESIWVASTDSSVHRWPAEGHNPQKMFQ RGGSFLAGNLSFSRARASLEGSTPIPVYREPTLK VPGIPAIVQHEILNNRRRVLTKDTAGSVKLWEITR GTVIEDYGEVSFEEKKKELFEMVSIPAWFTVDTR LGSLSIHLDTPQCFSAEMYSADLDIAGKPEDDKV NLARETLKGLLAHWLAKRKQRLGSQSSANGEVQ SVKDISTRSMTHSRIEVDGSSENDSMIYPPFEFSA VYPPSIITEGSQGGAWRKKVTGMDGTEDEKDLP WWVMDCVLNNRLPPRENTKCSFYLHPYEGSTIQ ILTQGKLSAPRILRILKVINYVVEKMVLDKSMEGG ATGDGTFATGLAGGPLQHSTTVGDGSLRSGLRP WQRPKPAIEILCNNQVLTAEMSLATVRTYVWKKP EDLVLHYRVMQGR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
848	sp_122700_knyk	-0.879098	#N/A	-0.852299	cytokinin oxidase 7 (CKX7); CONTAINS InterPro DOMAIN/s: Cytokinin dehydrogenase 1, FAD/cytokinin binding domain (InterPro:IPR015345), FAD-binding, type 2 (InterPro:IPR016166), FAD-linked oxidase-like, C-terminal (InterPro:IPR016164), FAD linked oxidase, N-terminal (InterPro:IPR006094); BEST Arabidopsis thaliana protein match is: cytokinin oxidase/dehydrogenase 1 (TAIR:AT2G41510.1); Has 8328 Blast hits to 8321 proteins in 1506 species: Archae - 208; Bacteria - 5322; Metazoa - 141; Fungi - 1495; Plants - 573; Viruses - 0; Other Eukaryotes - 589 (source: NCBI BLink).	MIAFIEHGFHHENDAESRQNDAVETSDQGDPAL WKEALKEIQGTVDYDSIFTGLAGKDFGGINRTKP VAVIRPAGSSDVAAVIRAVQTTPNLTVAARGNGH SINGQAMSDKGLVIDMRAMGDHPSLNVVVASSA EEDWSNSLSYADVSGGALWEEVLKRCVLGYGL APRSWTDYLALTVGGTLSNAGVSGQAFRYGPQT SNVTELEVVTGNGETVICSEFVYPDLFFAVLGGL GQFGVITRARIPLQKAPDMVRWIRLVYAEFDDFT RDSEFLVSQSKENSFDYLEGFVFVNSDDPINGW PSVTLDPDQWFDSDYIPPGSGSVLYCLEVALHFQ HHHHPTHVDSEVERLIQRLRFVKGLRFEVDLSYM EFLLRVKKAEQQATASGIWEAPHPWLNLFVSKR DIVDFNRTVFNKSLKGGIGGPMLVYPLLRSRWDS RTSVVIPEGEIFYLVALLRFTPPYPSGPPVVELVK ENMKIIEHCTKKGYDFKLYLPHYETQEKWKQHFG SQWPRFAERKAKFDPMAILAPGQKIFSRTHKTR QLL
849	sp_124640_jstu	#N/A	#N/A	-0.878261	tobamovirus multiplication protein 3 (TOM3); INVOLVED IN: viral replication complex formation and maintenance; LOCATED IN: endomembrane system; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1084 (InterPro:IPR009457); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF1084) (TAIR:AT1G14530.2); Has 262 Blast hits to 261 proteins in 45 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 200; Viruses - 0; Other Eukaryotes - 62 (source: NCBI BLink).	MAETLDPSLAVTAFLLKDASNWWNNINDSLLWQ DRIFHVLSILYGIVSVVALVQLVRIQLRVPEYGWT TQKVFHFLNFLVNGVRCVVFVFRRDVQLLQPEIV QHVLLDLPSLAFFTTYALLVLFWAEIYYQARAVST DGLRPSFFTINAVVYIIQIALWLILWWKPIPVLVILS KMFFAGVSLFAALGFLLYGGRLFLMLQRFPVESK GRRKKLQEVGYVTTICFTCFLVRCVMMCFNAFD KAADLDVLDHPVLNFMYYLLVEIVPSTLVLFILRKL PPKRGITQYHQIR
850	sp_124780_hwdq	#N/A	-0.865578	#N/A	uncharacterized protein	MNTKQRRNSNNKDNMEKQQNARNNIHYHILIFLR GSGTGRGSNINVFSSSMVGSNDDTYPRGYRKD HAIG
851	sp_124840_fjzy	-0.933029	-0.98773	-0.964288	PLANT HOMOLOGOUS TO PARAFIBROMIN (PHP); CONTAINS InterPro DOMAIN/s: RNA polymerase II accessory factor, Cdc73 (InterPro:IPR007852); Has 502 Blast hits to 395 proteins in 198 species: Archae - 0; Bacteria - 0; Metazoa - 257; Fungi - 127; Plants - 42; Viruses - 0; Other Eukaryotes - 76 (source: NCBI BLink).	MDPLSALRDYTIRGDLDKIHRVGDEFRFSTQYSF PCKTETAYRSKQGNLYTLETLVFCIKNRNCKQPE YIKNATNQGIPTVTYVDRKPVLDYLDGHVSSSDAI EFVLTTATDVEEYRPDDPQFFSSLPNNNGVSSLG IEDDSEIRVSEEGGGGGEGVDLMGLIRQSEKPLR DRNAVLQVRNRDFYGIYVGALRREQEKPQQRKD ANVGGNKGRPLGGEKDEGFHSGPKAKMLKGGG KIGEGVPIILVPSASSTLITIYNVKEFLEDGVFISTD VKMKQMKGPKPECVTVQKKFSRDKDKVMTAYE VRDKPSTLKADDWDRVVAVFVLGKEWQFKDWP FKDHVEIFNKIIGFYMRFEDDSIESAKVVKQWNVK IISISKNKRHQDRAAALEVWGKLEEFVARSRS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
852	sp_126050_udqo	-0.386583	-0.957039	-0.920951	Small nuclear ribonucleoprotein family protein; FUNCTIONS IN: molecular_function unknown; LOCATED IN: nucleolus, nucleus, small nucleolar ribonucleoprotein complex; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Like-Sm ribonucleoprotein (LSM) domain (InterPro:IPR001163), Like-Sm ribonucleoprotein (LSM) domain, eukaryotic/archaea-type (InterPro:IPR006649), Like-Sm ribonucleoprotein (LSM)-related domain (InterPro:IPR010920); BEST Arabidopsis thaliana protein match is: probable small nuclear ribonucleoprotein G (TAIR:AT2G23930.1); Has 1287 Blast hits to 1287 proteins in 269 species: Archae - 120; Bacteria - 0; Metazoa - 486; Fungi - 285; Plants - 213; Viruses - 0; Other Eukaryotes - 183 (source: NCBI BLink).	MVVGTLRGFDQFMNLVVDNTVEVNGDNTSDIGM VVIRGNSVVTVEALEPVSKMMQ
853	sp_127020_qwyy	#N/A	#N/A	-0.855383	ATB2; FUNCTIONS IN: oxidoreductase activity; INVOLVED IN: response to cadmium ion; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Aldo/keto reductase (InterPro:IPR001395), Aldo/keto reductase subgroup (InterPro:IPR020471); BEST Arabidopsis thaliana protein match is: NAD(P)-linked oxidoreductase superfamily protein (TAIR:AT1G60730.1); Has 30719 Blast hits to 30695 proteins in 2595 species: Archae - 650; Bacteria - 20319; Metazoa - 1822; Fungi - 2308; Plants - 1286; Viruses - 0; Other Eukaryotes - 4334 (source: NCBI BLink).	MAGVRRIKLGSQGLEVSAQGLGCMGMSAFYGP PKAEEDMIPLIQYAISSGITFLDTSDFYGPHTNELL LGKASTGGWREKVELATKFGVKFTDGKKEILGD PAYVRAACEGSLKRLDIDCIDLYYQHRIDTKLPIEV TMGELKKLVEEGKIKYIGVSEASASTIRRAHAVHP ITAVQMEWSLWTRDVEEDIVPTCRELGIGIVAYS PLGKGFFSSGPKMTENFTKEDFRQHLPRFKPENI EHNKHIFERVSEIATRKGCTPGQLALAWSITKEM MCVPSLEPLRLRTSTRTLEPCL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
854	sp_127340_gjsu	-0.9844255	-0.8765501	-0.9883734	BIG (BIG); FUNCTIONS IN: ubiquitin-protein ligase activity, binding, zinc ion binding; INVOLVED IN: response to auxin stimulus, indeterminate inflorescence morphogenesis, photomorphogenesis, root development, auxin polar transport; LOCATED IN: membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat-like-containing domain (InterPro:IPR011046), Zinc finger, N-recognin, metazoa (InterPro:IPR013993), Armadillo-type fold (InterPro:IPR016024), Zinc finger, ZZ-type (InterPro:IPR000433), Zinc finger, N-recognin (InterPro:IPR003126); Has 1182 Blast hits to 997 proteins in 123 species: Archae - 0; Bacteria - 4; Metazoa - 768; Fungi - 44; Plants - 189; Viruses - 0; Other Eukaryotes - 177 (source: NCBI BLink).	MTLASDMQLDNVTQSPTDSTFSQDFNKTVSSSQ HWSVLRLGCIHRLIMFCKQLVQLPEAFDVKTAEL SFCRKISLSLNILKLLRNLARSVPYVEVNSELLQSI SVFADMLPSLFRPSLDFISSNIASENYYEGVVLNIL QEFLHLVQLLSCGRSVFQNVRACLMASILDLLDS SVWRYNKSANNLKPPLVYFPRCVLYFVKLIGDIK NQGDEKQVEKDLGADLSFDIADPKARCVCHVRS ENILVKSYTAEELLEIIFPKSIHWVDNLLHLLYLLH TEGVKLQPKLEKSCSGAKSSSICDMENTGCHDD EALFGDLFSEGGRSVGSNDGYEQAPSVNPLTGF YNTIIEAADLSVFLKSCIFSPEWCAPVYEDACKK LNRNHIDSLLLMLGCQICYSEGFDGTTVVQEQRR IEHIQEICFEMLHGLLKRHAFSDLLEEHLVLQILKV ENGSVYYNDQTLVLLAQTLVCRVGSAGSNLRKK VCQQFVNFIVEKEKDVRSNCSILSELLISLPNVFH MEIILLTFHLSSESEKVALAELIFSNIKTLASSLQVS SSTQLSCWALVVSRLIVIFRHMMLNMRACPTSLL RDFRSKLNAARLVGSHHLHLNDPVTSWSSVVVD NVMSLFIKEEPLLGNLLNHLVDFESVPVRLFSSDA ANVNSSLSWNDIYTVFSPILELWKGKKAATIEDLV LERYMFLLCWDIPAAGLSPNHMSSAWSANCNLD PSNVEQFILFSHLLLGHCDLVKSHVDVVDIVIKFL QHLCSVSMPADIHEFGWDMMRNGQWLSLVLSF LHTGFGSYCVKNSVPGANFVWTENAFDAEFVSL AEGSISMILKTNKVPVILKSLSSLLAKYIRTYQRAF LFTISENQSHAQIFSPLLLLKHTEFDECMKDEILVK NGLVHAGLESVDDLLPKIDNVINKGGSHLLGVISS ESMLHGFPTHPRTSSGALVSSILSIRGIICLLDLIFK IRGTRGDISVETDVLHEILDSVMTVKYDRVLKNLS GKCRDVCRGCTGAEGPDYSLLFLIKQIEGFLSDIN VRGNVNHVKVEQLVVKAVDIMDSLRRDTLGIDVF RFYLGAEDVPQHVLDLYGSQRGDLLVLIDSLDSC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
855	sp_127930_wgty	-0.8770671	-0.8652352	-0.8767703	vernalization5/VIN3-like (VEL1); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1423, plant (InterPro:IPR004082), Fibronectin, type III-like fold (InterPro:IPR008957), Fibronectin, type III (InterPro:IPR003961); BEST Arabidopsis thaliana protein match is: Fibronectin type III domain-containing protein (TAIR:AT5G57380.1); Has 270 Blast hits to 250 proteins in 39 species: Archae - 0; Bacteria - 2; Metazoa - 8; Fungi - 14; Plants - 239; Viruses - 0; Other Eukaryotes - 7 (source: NCBI BLink).	MEERREIVYELAKWSHGAAEMLQSWSRQDILQIL CSEMGKERKYTGLTKLKIVEHLLLKVVSEKNSVNH DNEVRGQQSSPPITQKVSKKPRKPENPSHLLVS ATNISSCNSDSDLVTATYCKNSACKASLNKEDVF CKRCTCCICQKYDDNKDPSLWLTCSTEPPFVGD SCNMSCHLECAIKHDRSGIARNGFVSQLDGSFYC ISCGKMNDLMSCWRKQLIIAKDTRRVDILCYRISL SQKLLAGTKKYLKACDIVDDAIKLLAAEVGALTGV PVKMGRGIVNRLSSGQEVQRLCASALELFESTY HQTPLHLGNNVAVSRTIAPSIIRVESINATAVTVIL GPDNSSLSEKNVGYTLWHRKVGDRKKPEQPTC TLFPPQTRFIVSGLSPATEYIFKAISFDNTKELSSS EVRLTTNSTLERSQSPTTNCSTLSNPSSVEDET NNNAPCSNLNENNKAENYFNCSNDQNPSCASDE ENNATGKDSSDLSDTVKVEIKHSSDGIGSKDTKK TNTPCVRPETSLPITPSRQESSKEGTVKTGPPKS RNKDQDSRLDKKRKAERENGECVGDGSNSSGS NIEYYIKMIRWLECEGHIDKSFRQKFLTWYSIRAT PQETRIVKAFVDTLIDDPASLSEQLQDTFEEIISSK RSSAVPPGFCLKLWH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
856	sp_127950_nmud	0.8037739	#N/A	#N/A	GYF domain-containing protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: GYF (InterPro:IPR003169); BEST Arabidopsis thaliana protein match is: GYF domain-containing protein (TAIR:AT1G24300.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MASSDSRKPTLTPPNQIAKDITGPDNPIPLSFQW LLPKPGETKTGMSLGENRPSPYTGYGDRSDVTK SPGNGEEMHDNQKKRDVFRPTFLDAETGRRDR WRDEERDTNPSARKDRWREGEKEIGDGRKTDR WVDNTPTKNFGEARRAPSDKWADPSKENNYDQ RRESKWNTRWGPDDKEADKWTDAGKHGDMPQ DKGQSQLANHGKDEREGDNHRPWRPSSLQNRV KAETLHHHPQMTSKEGPTFGHGRGRGDNATFS VGRGRANFSAGSISNYSVHPHSGFNSDRGDNGH EEPSPFSYNRTKLLDIYRLTEIRSSNKILDGLKQVS SLTQEEPLGPLALCAPTPEESAIMKGIDKGDIISS GAPQNSKEGSTGRSPNEYAQSRRTKHGSRDDV SLPPDEDNDKVDHAKGTYSKISELSPREKQMNV YGSNVIMDGNENPQIYAEKYRVPAARDDESSLKK GDEVSDNRDPGIKGRVYSQSGTPWRSPSLEEHA STHSRDWFNSPAGGGIRSKSSDISWSQAPKDNS SVLPDPNAHNELKWQVGEDSAARKHAVVTVDR EQETRKLLHTSPDDMLLYYKDPQGQIQGPFSGS DIIGWFEAGYFGIDLRVRPVNAPNDVAFASLGDV MPHLRAKARPPPGFNTAKQSEVPDVSSRSSSSG FGKVLPGPSGVDLLGNEHRHTHGSSTDAENKFL ESLMAGEITTSHLERFASSEGLQRFVGQNSSGM SPSVAESGDALNLLAQRMNLERQKSLTQSYPYW ASREGAPVGPDFVRNSTIPQANPLSTVGDIPRQS LPQNVEMMSILQGLSDRSSGGVNNGLGGWPKF SVQGGIDPFKENVDLHSGQNFPQMAYGMQQHK QQPQNHPSMTNLLAQTLDNPAIQSDKLLSAGLTH DPQALNLLQQQYLMQLHSQQAVPGPQLSILDKFL LLKQQQKQEEQQQMLRQQQQQQLLSQVLSEQQ LVQRFGDPAFGQLQTSASPIGNAPIEQSAPQLSP SLFQANSQVPVSSMPDDRNVSIANMPPKFSQGL GHQVPADGSSLHLPHQMLGTAQKNWPSMSTHA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
857	sp_128300_onxz	-0.3856848	#N/A	#N/A	HEAT SHOCK PROTEIN 81.4 (Hsp81.4); FUNCTIONS IN: unfolded protein binding, ATP binding; INVOLVED IN: protein folding, response to stress; LOCATED IN: cytosol, apoplast, cell wall, nucleus, plasma membrane; EXPRESSED IN: male gametophyte, cultured cell, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Chaperone protein htpG (InterPro:IPR001404), Heat shock protein Hsp90, conserved site (InterPro:IPR019805), Heat shock protein Hsp90, C-terminal (InterPro:IPR020576), Heat shock protein Hsp90, N-terminal (InterPro:IPR020575), Ribosomal protein S5 domain 2-type fold (InterPro:IPR020568), ATPase-like, ATP-binding domain (InterPro:IPR003594); BEST Arabidopsis thaliana protein match is: heat shock protein 81-3 (TAIR:AT5G56010.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MVEAETFAFQAEINQLLSLIINTFYSNKEIFLRELIS NSSDALDKIRFESLTDKSKLDAQPELFIHIVPDKT NNSLSIIDSGIGMTKADLVNNLGTIARSGTKEFME ALSAGADVSMIGQFGVGFYSAYLVAEKVIVTTKH NDDEQYVWESQAGGSFTVTRDPGESIGRGTKIT LFLKEDQLEYLEERRLKDLIKKHSEFISYPISLWTE KTIEKEISDDEDEEKKDEECGKVEDVDEEKEKEE KKKKKIKEVSNEWTVVNKQKPIWMRKPEEITKEE YAAFYKSLTNDWEEHLAVKHFSVEGQLEFKAVLF VPKRAPFDLFDTKKKPNNIKLYVRRVFIMDNCEDL IPEYLSFVKGIVDSEDLPLNISREMLQQNKILKVIR KNLVKKCVELFFEIAENKEDYNKFYEAFSKNLKLG IHEDSTNRTKFAELLRYHSTKSGEEMTSLKDYVT RMKESQSEIYYITGESKKAVENSPFLEKLRKKGY EVLYMVDAIDEYAVGQLKEFEGKKLVSATKEGLK LEESEDEKKKAEEQKEKFEGLCKVIKDVLGDKVE KVIVSDRVVDSPCCLVTGEYGWSANMERIMKAQ ALRDSSMAGYMSSKKTMEINPENPIMDELRKRA DADKNDKSVKDLVLLLFETALLTSGFSLDEPNTF GNRIHRMLKLGLSIDDDEGEADVDMPALEEVDTE GSKMEEVD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
858	sp_128340_jksw	-0.825707	-0.88943	-0.903387	global transcription factor group A2 (GTA2); FUNCTIONS IN: transcription elongation regulator activity, structural constituent of ribosome, sequence-specific DNA binding transcription factor activity; INVOLVED IN: translation, regulation of transcription from RNA polymerase II promoter, positive regulation of RNA elongation from RNA polymerase II promoter; LOCATED IN: ribosome, intracellular; EXPRESSED IN: guard cell; CONTAINS InterPro DOMAIN/s: Translation protein SH3-like (InterPro:IPR008991), Transcription elongation factor Spt5 (InterPro:IPR017071), Transcription antitermination protein, NusG, N-terminal (InterPro:IPR006645), KOW (InterPro:IPR005824), Ribosomal protein L24/L26, conserved site (InterPro:IPR005825), Transcription elongation factor Spt5, NGN domain (InterPro:IPR005100); BEST Arabidopsis thaliana protein match is: Transcription elongation factor Spt5 (TAIR:AT2G34210.1); Has 14630 Blast hits to 9620 proteins in 607 species: Archae - 121; Bacteria - 647; Metazoa - 6069; Fungi - 2592; Plants - 1061; Viruses - 307; Other Eukaryotes - 3833 (source: NCBI BLink).	MGRGDDYEDDDDIEPEDGDEYGEDLYEEEPEEE DEDDDYGGRGGGGGGKGGKGSVKKRQRSDFIDD VAEEDDDEEEDEEVYARGRRHKKSRGGGFIDDI AAEVDDDEEEEEDGEDDFIVDTTTDVVEDERRM HRRPLLSREEQEEDVEDLERRIQERYARSSQSE YDEETTDVEQQALLPSVRDPKLWMVKCAIGRER EVAVCLMQKCIDKGSELQIRSVVALDHLKNFVYV EADKEAHVREAVKGMRNIYPSKILLVPIKEMTDVL SVETKAIDLSRDTWVRMKTGTYKGDLAKVLDVD NVRQRVTVKLIPRIDLQVLANKLEGRDVAKKKAF VPPPRFMNTEEARDLHIRVERRRDSMTGDYYENI NGMMFKDGFLIKSVSIKSISTQNIQPTFDELEKFR HPGEGDGDIGSLSTLFATRKKGHFMKGDAVIVVK GDLKNLKGWVEKVEDGNVHIRPNEKGLPKTLAIN EKDLCKFFEPGNHVKVSGSHEGATGMVVKVDF HVLIIISDTTKEDIRVFADHVVESSEVTTGVTKIGG FELHDLVLLDNATFGVIIRVESEAFQVLKGVPDRP EVVLVRLREIIRSKIDKKLFAQDRSKNPVAVKDVVK VLEGPCKGKQGPVEHIYRGVLFIFDRHHLEHAGFI PSRTPMHPYMTPMRDARATPIHDGMRTPMRDR CAKSQSCLVVGGSRGNREKMGDPLSYLRTPQRI PQSPGRLLPPRGPQSDFGGRGRGGGRGHDSLIG TTIKIRLGPYKGYRGRVKDIKGTTVRVELESQMKE VTVERSYISDNVTVTTPYRDTSRYGMGSETPMH AWNPYAPMSPARENWEDANPASWGSSPQYRP GSPPSRTYDAPTPVGWASTPGGNYNDPETPRD GNAPSPYLPSTPGTQPTTPSSVSYLPGTPGGQP TTPGGGGLDMMSPMIGGENEGLWLLPDILVNVIK SGDDSMVGVVQEVLPDGSCRVALGTSGNGEVIT ALPNEIDLIPPRKADKIKIMGGAQRGFTGKLIGVD GADGIVKVDESLDYKILDMGILAKLA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
859	sp_128920_jkap	-0.987017	-0.975312	-0.970757	Pre-mRNA-splicing factor 3; CONTAINS InterPro DOMAIN/s: Pre-mRNA-splicing factor 3 (InterPro:IPR013881); BEST Arabidopsis thaliana protein match is: Pre-mRNA-splicing factor 3 (TAIR:AT3G55930.1); Has 22923 Blast hits to 12760 proteins in 707 species: Archae - 18; Bacteria - 1085; Metazoa - 11525; Fungi - 3053; Plants - 1999; Viruses - 93; Other Eukaryotes - 5150 (source: NCBI BLink).	MEKATEKEKSSRKSREEERETTEHKHRHHRSRHR HRYEDEEDDDEHRRSRREKSHDDEHHRHHHK SREYSVDEQGDEEEDARDSRRGREGSSRDKDKD RKKRKDREVSEERIDDKRAKVSDVKKERRRFDD DDNGRIVKDERKEKRREDDGNGRVKIEEAKTKK EDGGRGDGSGEIVDNNNGRKRDRRRREDGDVK KESVAEPTGRSGSFPNDGAEKSLNVSSRFSSET FPANNHTATKVSSISTTHENKGVNITGSHAVTGK SSTDGTPSTAGKTGNISLDALAKAKKALQMQKEL AAKMKKIPLLNRGTSSSNAVNAQVGSKEEIKPPS AAGIQQTPSTSTFPPAAASAKTASMIDPAKYEAV KRAQELAAKMGFRQDTEFAPLINLFPGQLPLDTT ALQKPAKAPVLRLDAHGREVDEDGNVVYISKPSS LSTLKVNINKQKKDAFQIHKPELEVDPDKNPHFD PRMGIDKTKLLRPKRMNFLFVEEGKWSKDAEIIK LKSQFGEAQAKELKAKQAQLARAKAEPNINPNLI EVSERVITKKEPKESIPEVEWWDVPLLRSGDYAD IADGSSIEDRVKMEKITIYLEHPRPIEPPAEPAPPQ PQPLKLTKKEQKKLRTQRRLAKEKDRQEMIRQG LIEPPKPKVKKMSNIMRVLASEATQDPTKMEMEVR TAAAEREQAHVDRNIARKLTPAERREKKERKLFE DPNTLETIVSVYKINDFSHPQNRFKVDINAQENRL TGCAVISEGFSVVVVEGGSKSIKRYGKLMLRRIN WLAKVLNEDGEEDEDEEKPPNKCSLVWQGSVA KPSFNRFFIHDCRTEAAARKVFSDAGVGHYWEL AANFVDDQI
860	sp_128970_nfkt	-0.872043	-0.850373	#N/A	homeobox 51 (HB51); FUNCTIONS IN: sequence-specific DNA binding, DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: in 6 processes; LOCATED IN: nucleus; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Homeobox, conserved site (InterPro:IPR017970), Homeobox (InterPro:IPR001356), Homeodomain-like (InterPro:IPR009057), Helix-turn-helix motif, lambda-like repressor (InterPro:IPR000047), Homeodomain-related (InterPro:IPR012287); BEST Arabidopsis thaliana protein match is: homeobox protein 22 (TAIR:AT2G36610.1); Has 11020 Blast hits to 10992 proteins in 578 species: Archae - 2; Bacteria - 0; Metazoa - 8622; Fungi - 156; Plants - 2028; Viruses - 0; Other Eukaryotes - 212 (source: NCBI BLink).	MKRRLTTKQLESLESSFEEERKLDPDRKMKLAR DLALQPRQVAVWFQNRRARWKAKKLQHLYDAL KLDFELVSKEKHKLQQEVERLKSIMVMEEATKKQ VYSSYTEASCEETVESTSAAAANQYTGRNSGSN NGADDIVLN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
861	sp_129010_fpak	-0.838837	-0.836402	-0.834238	Ubiquitin-like superfamily protein; CONTAINS InterPro DOMAIN/s: Ubiquitin subgroup (InterPro:IPR019956), Ubiquitin (InterPro:IPR000626), Ubiquitin supergroup (InterPro:IPR019955); BEST Arabidopsis thaliana protein match is: Ubiquitin-like superfamily protein (TAIR:AT5G42220.1); Has 11489 Blast hits to 5746 proteins in 742 species: Archae - 0; Bacteria - 401; Metazoa - 4954; Fungi - 1390; Plants - 2430; Viruses - 186; Other Eukaryotes - 2128 (source: NCBI BLink).	MAEQQSTEGASTSSVPVVADSNMELNIKTLDSQ MYNFHVDKNMSVSSFKGMVADKVGVPVGQQRLI FRGKVLKDEHLLSEYHVENGDTLHLVERQPNQP QTSSGTGSTDASGNIGTQGMDGNAGPPRRIGQI SHSVVLGTLNVGDQGEGGVPDLSRVIGAVLNSL GVGNQTTNAAFGAPNTAPANSSASVAQANDTEQ GRGGSGGNQAQFGQPFLSQPFHVPLAGISVPPP SLNMPISDSLNTLSEFMSRMEGVLSQNGYQPSP PPSGTADQPGTELPSNFRGMATPEALSIVLRHAE RLLSGPAVAALSHIAGRMEQEAGSTDVSVRSQIQ SESVQVGLVMQHLGALFLELGRTILMLQMGRSP AEALVNAGHAVYISPSGPNPLMVQPFFHQTSPLF NAPAVSATNTGTSVPVGIIGSAPRNINIHIHAGTPL APLASAVGARTAGGEGTQGEHGGVTAASDSGG RVISVRNVVAASVPSHTATIAVSPVTTASQPVGV VSTSQPPDAASLSSVISHVNSQIRSLLENMRGDN QNPSGQQDNPVIPNSSAGSGGGGREAHESHTSSL PRQQAEGDKTDRRENIQSDSSQSNEESPSCMS GDPAAKSIGPKEVPLGLGGGLQPKRRGRQSQM QPVAGEAGPSSSFDQSQQARISGQQILQSLASR NTAGRSSLTNPGSGQTGQGIEQSVAGTQVTQG SDGQFDITGSMSQLLQSPALDGLLSGVAQQAGV GSPNVLRNMLQQFTQNPAMRNTVNQLAQQVDR EEFRNMFAGMGTGGDGTPDFSSMIQQMMPVVS QVLGGGSIQNEQLRNMVPQLMPQSDDRSIEDSS LGVVSQEVVRRIALHNSSEDIFRSVVQNAVENTE NTETATREFIDDLCSNEDLIDEFVEMLRCDLRRRL RDNFV
862	sp_129111z4911	#N/A	-0.865181	#N/A	uncharacterized protein	#N/A
863	sp_129120_exgt	-0.853244	#N/A	#N/A	ubiquitin conjugating enzyme 8 (UBC8); CONTAINS InterPro DOMAIN/s: Ubiquitin-conjugating enzyme/FWD-like (InterPro:IPR016135), Ubiquitin-conjugating enzyme, E2 (InterPro:IPR000608); BEST Arabidopsis thaliana protein match is: ubiquitin-conjugating enzyme 28 (TAIR:AT1G64230.2).	MASKRILKELKDLQKDPPTSCSAGPVAEDMFHW QATIMGPADSPYAGGVFLVTIHFPPDYPFKPPKV AFRTKVFHPNINSNGSICLDILKEQWSPALTISKVL LSICSLLTDPNPDDPLVPEIAHMYKTDRSKYEATA RSWTQKYAMG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
864	sp_129360_krgp	-0.9387225	-0.933359	-0.8736379	Sec14p-like phosphatidylinositol transfer family protein; FUNCTIONS IN: phosphatidylinositol transporter activity, transporter activity; INVOLVED IN: transport; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Cellular retinaldehyde-binding/triple function, C-terminal (InterPro:IPR001251), Cellular retinaldehyde-binding/triple function, N-terminal (InterPro:IPR008273), Cellular retinaldehyde binding/alpha-tocopherol transport (InterPro:IPR001071), Phosphatidylinositol transfer protein-like, N-terminal (InterPro:IPR011074); BEST Arabidopsis thaliana protein match is: Sec14p-like phosphatidylinositol transfer family protein (TAIR:AT2G21520.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MQRCQILNKGPWKDPEIAKLIKSGATKLAISNSEE RVIADNKMVKNNDVSTAESGSEVMHLFSSRAPG AAALKWKLMPVDEEI

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
865	sp_129740_zedu	-0.9726824	-0.9511729	-0.9496888	THO5; CONTAINS InterPro DOMAIN/s: THO complex, subunit 5 (InterPro:IPR019163); BEST Arabidopsis thaliana protein match is: THO complex, subunit 5 (TAIR:AT1G45233.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MDVKMEDAAAAAAAAAAATATATSATSEDIIVLS KTAEKVEKSPYDMLKESKISVEEIVAKMLSLKKDD KAKPELRELVTQMFLHFVSLRQANRNILIEEDRVK AETERVKAPVDSTSLQLHNLMYEKSHYLKAIKAC KDFKSKYPDIQLVTEAEFFSQAPEEFKGIAISNET AHNVMLKRLNYELFQRKELCKFREKLDQKKKSLL ETIANRKKFLSSLPSHLKALKKASLPAQNQLGLLH TKKQKQLHAAELLPPPLYIIYSQFMAQKEAFGENI DLEITGSLKDAQVFVRRQTYKDTGTSSMQESSRI DDDVVDEEDDGQRRRKRPKKVASKENVDTAGIY QAHPLRIILHVYDNEVCEPKSARLVTLKFEYLFKL NVVCVGVEGSDQGSENNILCNLFPDDKGLELPQ QSAKLLLDDGAEFDERRTMRPYKWAQHLAGIDF LPEVSPSSTSNEAGKFETGKDAVVSGLSLYRHQ NRVQTVVQRVRLRTKAQLALSEQLDSLTALKWP TLICQSVPWAAHEHLCFLHRWSHVGSSLNQDGS LSRVDVEQDHDAFDANMAGRASSSKEEGIMTED GELPSLTPATVVTDEKSSAYLNQGSGVKHSKRM ALISKSIAVPVKTARNQSFGRYEEDTDLMMDSES EVDDTTNNDKEIECAVDEGNSWVDYGIKEFNLVL TRKRDAGEGNVDLEVKIRISKEYPLRPPLFKLSLL SQYSGGNSSPIDHSVWYNELRAMEAEVNVHILK MLPPNEENNILSHQVHCLAMIFDLYLNEASQPSQ GRKCKDSDIGSVSGGLKARSFRGRDRRNMIPWK DMELTPGYPF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
866	sp_130120_mkrp	0.8953815	0.9423615	0.9453889	secret agent (SEC); CONTAINS InterPro DOMAIN/s: Tetratricopeptide TPR-1 (InterPro:IPR001440), Tetratricopeptide-like helical (InterPro:IPR011990), Tetratricopeptide repeat-containing (InterPro:IPR013026), Tetratricopeptide repeat (InterPro:IPR019734); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT3G11540.1); Has 114710 Blast hits to 35863 proteins in 2480 species: Archae - 3197; Bacteria - 54774; Metazoa - 16661; Fungi - 4580; Plants - 3873; Viruses - 0; Other Eukaryotes - 31625 (source: NCBI BLink).	MISLRTESRQQCHHHKNSDLQLSKSSFDLKSDVK FNHFKGTESREDVKEDVLLGAAHQLYKAGDYKK ALEHSNTVYERNPLRTDNLLLLGAIHYQLHDFDM CIARNEEAIQIDPHFAECFGNMANAWKEKGNVDL AIRYYLIAIELRPNFADAWSNLASAYMRNGRLIEA AQCCRQALALNPLLVDAHSNLGNLMKAQGLVQE AYSCYLEAIRIQPTFAVAWSNLAGLFMESGDLNR ALQYYKEAVKLKPSFPDAYLNLGNVYKALGMSQ EAVTCYQHALQARPNYAMAFGNLASTYYEQGQL DLAILHYKEAINCDSGFLEAYNNLGNALKDVGRV DEAIQCYNQCLSLQPHHPQALTNLGNIYMEWNM VAVAATYYKATLSVTSGLSAPFNNLAVIYKQQGN YADAISCYNEVLRIDPLAADGLVNRGNTYKELGR VSEALQDYINAINIRPTMAEAHANLASAYKDSGHV EAAIKSYKQALELRPDFPEATCNLLHTLQCVCSW ENKDKMFSEVEGIIKRQVKMSVLPSVQPFHAIAY PLDSMLALEISRKYAQHCSVVASRYALSPFKHPQ PITIKSEGGFQRLRVGYVSSDFGNHPLSHLMGSV FGMHNRENVEVFCYALSPNDGTEWRQRTQSEA EHFVDVSAMSSDMIAKLINQDNIQILVNLNGYTKG ARNEIFAMQPAPIQVSYMGFPGTMGASYIDYLVT DEFVSPFRSAHIYSEKLVHLPHCYFVNDYKQKNR DLLEPNCHPKRSDYGLPEDKFLFACFNQLYKMD PEIFTTWCNILKRVPNSALWLLKFPAAGESRLRA YAAAQGVGQDQIIFTDVAAKNEHIRRSALADLFLD TPLCNGHTTGTDILWAGLPMITLPLEKMATRVAG SLCLATGLGEEMIVHSIEDYEERAVALALNPSKLE SLRNQLKAVRLTCPLFDTARWVKNLERAFYKMW NLYCTDKHPQPFKVTENDVDFFLDR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
867	sp_130680_kgmh	-0.8867393	#N/A	-0.8717736	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity, ATP binding, nucleic acid binding, ATP-dependent helicase activity; LOCATED IN: cellular_component unknown; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEA(D/H)-box RNA helicase family protein (TAIR:AT5G60990.1); Has 35246 Blast hits to 34671 proteins in 2939 species: Archae - 524; Bacteria - 16250; Metazoa - 5646; Fungi - 4238; Plants - 2336; Viruses - 4; Other Eukaryotes - 6248 (source: NCBI BLink).	MANSSNTTQKEEIELDDEENQSFEDLSLDPRLLR ALIKKGIKKPTPIQRVAIPLILEGKDVVARAKTGSG KTFAYLLPMLQKLLLDSGSNKRLAPSSFILVPTRE LCHQVYSEVTSIIELCKAQLKAVQLIRGMTAAQMK QTMSGPPDILVSTPACVLTCLKAGILQPKAIQDSL SMLVLDEADLLLNYGHGDDLRSLTAHVPKRCQC LLMSATPSADVEKLKKLILHNPYILTLPEVGGNKD ELIPNNVNQSWVSCKSDHDKFIYILALLKYEFVAK KMLIFTNTIDMGFRLKLFLEKFAFRSAVLNAELPQ NSRLHMIEEFNAGLFDYLIATDIGQTGKDQEDSVK KPEGKNSRKRKKKPKLDAEFGVVRGIDFKNVLTVI NFEVPQTASGYVHRVGRTGRAFTTGASVSLVSP DEMEAFDEIKATLSDKESGKAYFINSFNSKGLDE AVKSLRYRAEDILRSVTSFCVRESRAQDLRNEILN SEKLKSHFEANPRDLDLLKHDKVLSKKEPPAHLR NIPDYLVDPATQEASNIVKLTRLAMGNNRKPSQR PSQKRKSKRKRDPLKTLSGEGSNQRNKFKRSN QK
868	sp_131650_jnii	-0.8844076	-0.8653831	#N/A	CONTAINS InterPro DOMAIN/s: Membrane protein,Tapt1/CMV receptor (InterPro:IPR008010); Has 447 Blast hits to 428 proteins in 176 species: Archae - 0; Bacteria - 0; Metazoa - 190; Fungi - 133; Plants - 49; Viruses - 0; Other Eukaryotes - 75 (source: NCBI BLink).	MASKSRGKQLSFEILSSTYAVNSDSIIDRSSNSDP LIQNNKSNRRRRKRKRKKSTPNSPAPHEFYTIFD NGGIDELEERESPSMNSHSYAASSFVVCEVREE SLGTVCQVSQGELRQRNVSSGGSGGSGGDLVPPVEI PEAEVKSRPNLINGGKLDKSESLDWKRFMAEED PMYPFPVENSPIKFFMEEMYNGNSIRSTTTRGTE KERERVYDTIFRLPWRCELLIDVGYFVCLNSFLSL LTIPMRALMTLWRFLTKRQPRTPSAAELSDFGC LLVLAVGVTTLGRTDISLIYHMIRGQGIVKLYVYYN VLEIFDKLGQSFGGDVFQTLFNSAEGLACSSPES SGYWLERYTTDLILAAAASIVHSFILLAQAITLSTCI VAHNNALLALLVSNNFAEIKSNVFKFRLSKDNVHGI VYSDSIERFHSAFVIFVLAQNIQEAEGPWFGNFL SNALLVYFSEMVIDIIKHSFIAKFNDIKPIVYSEFLE DLCKQTMNIETDGAKKTLTFVPVAPACVVIRVITP VYAARLPCSPLPWRYFWMICLTMLSYVVMLVSLK VMIGMGLQKHAAWYIKRCQKRKLHSD

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
869	sp_131930_cijt	#N/A	-0.862465	-0.868065	POZ/BTB containin G-protein 1 (POB1); CONTAINS InterPro DOMAIN/s: BTB/POZ (InterPro:IPR013069), BTB/Kelch-associated (InterPro:IPR011705), BTB/POZ fold (InterPro:IPR011333), Kelch related (InterPro:IPR013089), BTB/POZ-like (InterPro:IPR000210); BEST Arabidopsis thaliana protein match is: BTB/POZ/Kelch-associated protein (TAIR:AT2G46260.1); Has 3471 Blast hits to 3456 proteins in 95 species: Archae - 0; Bacteria - 0; Metazoa - 3227; Fungi - 0; Plants - 131; Viruses - 11; Other Eukaryotes - 102 (source: NCBI BLink).	MKNMSRDLFDPRTVMDSEYDRSAKEGNFAFAFN DSNFSDRVLRIEIMAGPSYSKSDADGCCSLFDW ARNRKRRREDLKKDVQIDNDTIEGEDVTVGPEEQ VLHQSDGDDVLACENDDEEAVAMIEELPSGDEG TNSNDSLWSMECSTIVSVKTLHISSPILAAKSPFF YKLFSNGMRESEQRHVTLRINASEEDALNELLNF MYNNTLSANTAPALLDVLMAADKFEVASCMRYC SRQLRNLPMNPESALLYLDLPSSVLMAEAVQPLT DAAKQYLAERYKDITKFQEEVMGLPIAGIEAILAS DDLQVASEDAVYDFVLKWSRSQYPKLEERREILA SHLARYIRFPYMTCRKLRKVLSCPDFDHEVSSKL VLEALFFKAEAPHRQRILAAEESASTNRRFLERA YKYRPVKVVEFELPRQQCVVYLDLKREECANLFP SGRVYSQAFHLGGQGFFLSAHCNMDQQSSFHC FGLFLGMQEKGSVTFAVDYEFAARSKPTEEFQS KYKGNYTFMGGKAVGYRNLFGIPWTSFMAEDSL YFINGILHLRAELTIRH
870	sp_132250_ugpw	#N/A	#N/A	-0.855372	SUPPRESSOR OF ACTIN 1B (ATSAC1B); CONTAINS InterPro DOMAIN/s: Synaptojanin, N-terminal (InterPro:IPR002013); BEST Arabidopsis thaliana protein match is: Phosphoinositide phosphatase family protein (TAIR:AT3G51460.1); Has 1740 Blast hits to 1632 proteins in 227 species: Archae - 0; Bacteria - 0; Metazoa - 569; Fungi - 574; Plants - 290; Viruses - 0; Other Eukaryotes - 307 (source: NCBI BLink).	MRLWEFPDEYIVEPTDGTGGSFLSVSRVDGSMK LLDELPQCSAERVPKIQTIYGVMGALKLLAGMYLF VITERESVGSYLGHPIFKVSSMKILPCDQSLKSSP VEQKRMETEFSSLLNAAVKTPGLFFSYDVNVTLS MQRLHDLGNEEKTIPLWRQADPRYCWNNYMME MLIDNKLDPYLLPILQGTFHHFQAGVGNDTIDVTLI ARRCTRRTGTRMWRRGADPDGYVANFVETEQI MQLNGYTASFVQVRGSMPFLWEQIVNLKYKPKF EIVNPEEAPRVAERHFLDLRKRYGAVLAVDLVNK NGGEGQLCKMYGSAMQKIVNEELRYLHFDFHHI CGHMHFDRLSILYDQISDFLNKNRYLLLNEKGEKI EEQSGVVRTNCIDCLDRTNVTQSMIGRKMLELQL RRIGVFGAEETISSHPKFDENFKILWANHGDNVSI QYSGTPALKGDFVRYGKRTFEGVFNDGRNALAR YYLNNFCDGTRQDAIDMVHGHLIASVSRDMTTQ PQKGGLEAIASVPVGLSLVLAGIVFAFMSLKQVG LDWRHLLFSLVWASLSIAIAIFVRANGRVFCNRPR LHKPRC
871	sp_132390_emhc	#N/A	-0.90287	#N/A	uncharacterized protein	MSYRRRGGLNGRGAIIRPPGYRSNGLTNTHSTG EDTPDQLPNYGHVAEETTLNHVRNSRYQAAASM ATTNTTMEATNTTRQLPNTVHQVLRQVPHQVPH TIPLHNADVGNNHETITNITDETSNPEKDEHVLHP DGLWFPHRAVVNKVSATSYKSYFKGPYCNWKM TPPQVQQRWWNAFKHEFSWDPSVAKLVKKGW KSKVSRRLTGIVSKEKKLGYQPSPSEVYYETHAD SGGCFVNAKAEAVWEKKLGYQPSPSEVYYETHA DSGGCFVNAKAEAVWVSIME

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
872	sp_132510_stum	#N/A	#N/A	-0.857303	decapping 5 (DCP5); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: negative regulation of translation, deadenylation-independent decapping of nuclear-transcribed mRNA, cytoplasmic mRNA processing body assembly; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: DFDF motif (InterPro:IPR019050), FFD/TFG box motif (InterPro:IPR019053); BEST Arabidopsis thaliana protein match is: decapping 5-like (TAIR:AT5G45330.1); Has 14874 Blast hits to 10327 proteins in 603 species: Archae - 12; Bacteria - 902; Metazoa - 6418; Fungi - 2978; Plants - 1226; Viruses - 261; Other Eukaryotes - 3077 (source: NCBI BLink).	MVDGPDGNLSISESVIDKNLKSWRPKLRSRLHPP SYIGSVISLTSKSEIRYEGILFSINT
873	sp_133880_nmnw	-0.31655	-0.966321	-0.977681	Heat shock protein 70 (Hsp 70) family protein; FUNCTIONS IN: ATP binding; LOCATED IN: cell wall, plasma membrane; EXPRESSED IN: male gametophyte, guard cell, cultured cell, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp70 (InterPro:IPR001023), Heat shock protein 70 (InterPro:IPR013126); BEST Arabidopsis thaliana protein match is: heat shock protein 91 (TAIR:AT1G79930.1); Has 26529 Blast hits to 25703 proteins in 4372 species: Archae - 153; Bacteria - 11749; Metazoa - 3974; Fungi - 1605; Plants - 1160; Viruses - 126; Other Eukaryotes - 7762 (source: NCBI BLink).	MSVVGFDFGNESCVVAVARQRGIDVVLNDESNR ETPSIVCFGEKQRFIGTAGAASAMMNPKTTITQIK RLIGRQFSDPELQRDIKSMPFKVTEGPDGFPLIEV RYLGETRTFTPTQLLGMLFSNLKGIAEKNLNAAV VDCCIGIPTYFTDLQRRAVIDAATIAGLHPLRLIHE TTATALAYGIYKTDLPENDPINIVFVDIGHSSLQVC IASFKKAQLKVLAHAFDRSLGGRDFDEALFQHFA AKFKEEYKIDVYQNARACLRLRSSCEKLKKVLSA NPEAPLNIECLMEEKDVRGFIKREEFEQISLPILER VKKPLEKALADAGLSVENIHAVELVGSASRIPSVL KILTEFFGKDPRRTMNASECVAKGCALQCAILSP TFKVKEFQVNESFPFPISLSWKGVAADGQNGAA EGQQTTVIPKGSPIPCVKAVTMLRSGTFTVDVQ YADVSELQAPAKISSYTIGPFQSSNGDRAKLKVR VRLSLHGIVSIESATLLEEEEVEVPVVKETAKMDT DDAPTETPSGEADVNMQDVKGGAEIPSAENGTP ESAEKSTSEGDDKPVQMDTDVKADALKKKVKKIT VPVSELVHGGLSQGDLQKAVEKEFELALQDRVM EETKDRKNAVEAYVYDMRNKLNDKYHDFVDPSE KEKFIARLQEVEDWLYEDGEDEIKSVYVAKLEEL KKQGDPIEERYKDSMERGNVINQLTHCFQSFRE AAASNDPKFDHIDIAEKQKVINACSEAELWLGEK KQQQDALPKYATPALRSADVVKKSEEVDRLCRPI MTKPKPTPPKPATPEPQQPQQAQSGEAPSPGG DTDSGANDSNETPDAEPMETEKPENAA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
874	sp_135390_aamj	-0.910836	-0.883341	-0.960288	SHORT ROOT (SHR); CONTAINS InterPro DOMAIN/s: Transcription factor GRAS (InterPro:IPR005202); BEST Arabidopsis thaliana protein match is: GRAS family transcription factor (TAIR:AT3G49950.1); Has 2470 Blast hits to 2429 proteins in 304 species: Archae - 0; Bacteria - 4; Metazoa - 9; Fungi - 15; Plants - 2427; Viruses - 0; Other Eukaryotes - 15 (source: NCBI BLink).	MPYSIENPLSLRSDHSLVPKRSTSFPLQLPGKWL MNQTIKKPLKNVHHRESSICPPSSSTDNKSCVG WVSSLLCECVKAISQGNPSKSQQLLWVLNEVAS AYGDCDQRLAYYFLQALFAKANNSGLQFYSTIKF VEEKNYCFDTYVKLMLKFQEVSPWTTFGHVASN GAILEALEGENNLHIIDLSNTLCTQWPMLLESLAT RNDETPHLRLTLVVPTSLEIIITKEISKKMEKFARL MGVPFKFHVISGLDCLGQIKIEDLCLQDGEAVVV NCIQALQRVHVDKRRDVIDMIRSIKPRVVTIVEEE VDLTTITSSDFGKCFEECLKFYKLYFEMLEESFPP ISNERLKLESKRSRNIVNALACENGEIGEEYWRP DKASQWSKRFLNQAFTPFHYNDEILGDVRALLR RYNTRWDLALPQQHDHQVGIHLKWKDEPVVWA SAWKPNLSASPPNPS
875	sp_135490_icfw	-0.928473	-0.836328	-0.834554	C2H2-like zinc finger protein; FUNCTIONS IN: zinc ion binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system, intracellular; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), Zinc finger, C2H2-type (InterPro:IPR007087); BEST Arabidopsis thaliana protein match is: zinc finger (C2H2 type) family protein (TAIR:AT5G40710.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVTIRSKKSIGFTCFLLFVMLTNSSILGISAAEPLH HQDQGLGENAESRDLKKEHEHIHEVHCSRERSR AAWEIVEKYLMHFVERERYEISRNCRLHPDNDLF RDQENHKIKEDVNEWQCGYCKKSFLSEEFIDRH FDSRHYNLLNFTGSKCLADLCGALHCDHVLGSK SPRAKCNPAAAARNRHLCESLADACFPVSQGPS ASRLHELFLQQFCYAHTCTGGIKPFSKGGRKETS VLYFAASVLTLMLLPLFYVLVCLHQSETRKDTQEL KRISRLGRKAKPM
876	sp_135520_mahn	#N/A	-0.830849	#N/A	uncharacterized protein	MFEQRRRASRGKKAAVAVAAVAATATATSAKKV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
877	sp_135530_zhgh	#N/A	-0.850331	#N/A	EMBRYO DEFECTIVE 2423 (EMB2423); CONTAINS InterPro DOMAIN/s: Telomere length regulation protein (InterPro:IPR019337); Has 246 Blast hits to 237 proteins in 108 species: Archae - 0; Bacteria - 0; Metazoa - 102; Fungi - 75; Plants - 52; Viruses - 0; Other Eukaryotes - 17 (source: NCBI BLink).	MATEEKLKKSEEQTQLLETVSATISAVSTAKYVD DVILALHSVAVLIFPIDTSSLSGVISEKYSDQLNQA ETTYTSERRNQVWEVFYKGVSFPTLSRFLIYDVA SNWLPCFPLTARKLVYDIFFVEGLVLEIVQALIPSL GHNVSVVRSNAERLVELCLLENEGVLQLARELG AYHESKDSSNSLVKLTVSRVAQLVASVPDKARLG APKSLSSHLFFKSVTIQLLDGAEEKSWELYDKEV LFHGTDVDGTFQFVAEMFARICRRGSTELLVNEV ILRIRRRIQRFLLSHKDTITENSFKSNSDFQFWFL MIEILRDAYVVEKMIEKILHQLAAECEHDTEAYWIL WILFHRAFQHQITIRAMFVENFLLSKVFPFRCLR WILQFAIFGCPFDDRKLKKGDNNHVLLEAAQRVV QMWSRREFVQSAPLEQQVVYTAVAGLLLENMS REDLDSTKDVMRSILEGVSCRLESPDELVRRMAS NIALVFSKVVDPKNPLYLDDSCKEGNVDWEFGS ASSNSALNSKDRTSDNLKNTKTIDKGVHMKEQN HGDTHERESKSNPGKKKPFGHKLVDPDEIVDPA SLNDELLSGEEVDDNASEDSEASSDSSLQPYDM TDDDTDLKKRISQLVDVVGALRKPDDADGVESAL DVAEKLVRASPDELRHIASDLVRTLVQVRCADTV LEEEEESAEEKRQKALVALLVMCPFESLNAINKLL FSPHLDVSQRIMILDVMTDAAEELANAKTIVRKNQ PKSLISSASEPQPWFLPSSKRPPGAGSWKEIPDT LPFSALNLTNRFEERELPSKPGVIKGKTRRWGHKP INLQANEINWSQNKFPLYAAAFMLPAMQGFDKKS HGVDLLGRDFLVLGKLIYMLGVCIRCTALHPEASA LAPSLLDMISSREICHHKEAFVRRSVLFAVSCSLK ALHPSFIASAIMEGNEEVSRGLEWVRTWALRIAE SDTDTECCAMAMACLQLHAEMSLQASRAFNTAE SSLQKKSIGLPSNMSKGVIRIPGSNSFLG
878	sp_135710_jlya	#N/A	#N/A	-0.8503737	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G57080.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MNGKFGGGKLPTGTPSMARSTVVVVLSLLGGAS VVHNIYKPDLTLPSPEEVSEDKQKNEKQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
879	sp_136540_cxet	-0.321235	-0.957386	-0.314974	ENDOPLASMIC RETICULUM MORPHOLOGY 2 (ERMO2); FUNCTIONS IN: transporter activity, zinc ion binding; INVOLVED IN: transport, ER body organization; LOCATED IN: COPII vesicle coat; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Sec23/Sec24, helical domain (InterPro:IPR006900), Sec23/Sec24 beta-sandwich (InterPro:IPR012990), Sec23/Sec24, trunk domain (InterPro:IPR006896), Zinc finger, Sec23/Sec24-type (InterPro:IPR006895), Gelsolin domain (InterPro:IPR007123); BEST Arabidopsis thaliana protein match is: clone eighty-four (TAIR:AT3G44340.1); Has 86818 Blast hits to 46504 proteins in 1607 species: Archae - 60; Bacteria - 12149; Metazoa - 40940; Fungi - 13838; Plants - 8598; Viruses - 2013; Other Eukaryotes - 9220 (source: NCBI BLink).	MRPFLSSGPVAGRPDGDGVRPPPPSAPGSAMP FASPRPMAGPPVPGFRPSPVVGFTDQIRPGPPP SHAPPTSNNFQQFSAPQYSGPPTNPPIRTPPVG PPNMNMASPMPPNMPPPIQRPGVPAPTPPLTFR PNSQVPPQPYAQAPPQPFAQGPPQPYAQQGPPQ PYAQGPPQPFAQGPPQPFNQGPPQPFAQGPPQ PYAQGPPQPYAQGPPQPYGQPAYAQSPSQPGM QGFVEDFSSLSLGAVPGSMDPGLDSKALPRPLE GDVEPTVFNDMTMNCSPKYMRLTTSAIPNSQSL ASRWHLPLGAVVSPLAEASEGEEVPILNFGSAGII RCRRCRTYINPYITFTDGGRKWRCNMCALLNDV PGDYVAPLDATGRRIDADQRPELIRGSIEFVAPTE YMVRAPMPPLYFFLIDVSVSAVRSGMLEVVAQTI KSCLDEIPGYPRTQIGFITFDSTIHFYNMKSTLTQP QMMVVSDVDDIFVPLPDDLLVNLSESRNVVESFL DSLPSMFQDTVNVESAFGPALKAAYMVMSQLGG KLLIFQNTLPSIGLGRLRLRGEDPRVYGTDKEASL RVPEDPFYKQMAAECTKYQIGINVYAFSDKYTDIA SLGTLSKYTGGQVYYYPSFRSTIHADKLRHELAR DLTRETAWEAVMFIIRCGKGVKFNSYHGNFMLRS TDLIALPAVDCDKAFAMQLGLEETLLTTQTVYFQ VALLYTASCGERRIRVHTAAAPVVSDLGEMYRRA DTGAIVSVLSRLAIEKTLSSKLEDARSYLQMKIVK SLKEYRNLYAVQHRLAGRMIYPESLKYLPSYGLA LSKSMPLRGGHADALLDERCAVGSTMMTLPVKS LLKFLYPSLIRVDEFLLKASSPPDADLKNVEKRLPL TAESLDARGVFILDDGLRFVIWIGRMFPPELAASL NIDFASDLSQVNGSNNEMSRRLMGLLKKLRESD PAYYQFCHLVRQGEQPKEGFYVLSNLIEDQTAGT NGYVDWIAQIHRQVQQNA
880	sp_137640_jigi	#N/A	#N/A	-0.642215	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 10 growth stages; Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MASFTILGALASAIIVSMAVLPSSSALVSSPDVNP PFPKAISDLKEAVVKGLGFQTDDFKITGFDMRDT LVGHSVAYEFDVEIDNKVLPVKLLEDVKRWEYVD LPIFRVEKPVKPGEENGLIERKMMEENMGPVLAP FQLAGPMELWIQDAKHMRLSLPHDVDAGELRKVI LAEGAVVTVKGAKSVSLRQPLELPLPLNQTQNGF ASGLLSLAQHLHHVSGGSKGAPLLSLRVVGPTSLA SPSPSSHSSSNRLKLKRLAPGLVELSSPSTNKPV EIATAIDIPSEVPTLLTPEQFTTLWPMSSLNGSNS NLLGLEKLLSSVLGPKANKKGSFRLLKADVSAQT YLKVGFKVEKQLQEGDGINLEGFPEWRTKPEKV KMHFEALAKVEGDRVIPEKVTQVHPVILEDTVAP SVLLGNVTMSKMPIFHQPSNPLTL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
881	sp_137760_ximr	-0.3191148	-0.967788	-0.3486978	RNA binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: mRNA processing; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677), Splicing factor PWI (InterPro:IPR002483); BEST Arabidopsis thaliana protein match is: RNA binding (RRM/RBD/RNP motifs) family protein (TAIR:AT2G24350.1); Has 3110 Blast hits to 3077 proteins in 726 species: Archae - 48; Bacteria - 800; Metazoa - 1243; Fungi - 266; Plants - 328; Viruses - 0; Other Eukaryotes - 425 (source: NCBI BLink).	MEGENRDGGSTFRVNFSDEGAAKLAALLTDTLK EYMGDDVDDTLVKYVIVLLGNGRLKEEVKEDLDV FLGDKSDSFVTWLWDHLASNLDLYVGSPQSCPN EVAKAKPELVERTSKDHSDHPVSESDRGKSSKT FRSRHNRDWKGLVTDNSQPPLIRSSVTNNLQNE EKTYPNALHGRQSVSPERPVHRKRYRPGERPRP KREISQATIAAPRRLLQFAVREAVATTRSSNSVSE PARKRLRSVVSTTTGNSTEEVYHRRIQSAVRLQN PMATAINAVAEASKDVRRVRTGNVFDRLSRGDG SSDKVPVFKELAVQDDEHDNYDHIHEKSRHRYV EKRENNGHHNVNASIVNSSTGLASDSASDNEGY DDVNVVGARGVDVSETGTSAKADDSLMVHYSVD QNTDGVTRQSRRKDQELPASAANTKHKIVNISVN VNTWKPHHYQAAREVPDVEIPKSTQETDLGASN SNGNPAADTQKLNQKTAGSSLPGAHSTGRPLED ADSRTIFVNNVHFAATKDTLSRHFNKFGEVLKVIIL TDAATGHPKGSAYVEFMRKEAADNALTLDGTSF MSRMVKVLKKSSVQQEANPVMMWPRITRGSPF AAPRFAQAPFPKVFPGTQRARPPVKPGMRSMQ WKRDTQQTSVESVASVPGNLNAGRGLTYIRTDA K
882	sp_138010_ajjs	#N/A	-0.859345	#N/A	NADH-ubiquinone oxidoreductase 24 kDa subunit, putative; FUNCTIONS IN: NADH dehydrogenase (ubiquinone) activity, zinc ion binding; INVOLVED IN: response to oxidative stress, mitochondrial electron transport, NADH to ubiquinone; LOCATED IN: mitochondrion, mitochondrial respiratory chain complex I, respiratory chain complex I; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), NADH-ubiquinone oxidoreductase, 24kDa subunit (InterPro:IPR002023), Thioredoxin-like fold (InterPro:IPR012336); Has 5564 Blast hits to 5564 proteins in 1535 species: Archae - 26; Bacteria - 3396; Metazoa - 195; Fungi - 114; Plants - 49; Viruses - 0; Other Eukaryotes - 1784 (source: NCBI BLink).	MFARRAAQRLLEIRQVFRQSPQASRSFSSALNFH LDTPDNNPNNQWKFSETNNQKVKEILSFYPTNY KQSAVIPLLDLAQQQHGGWLPVSAMNEVARIIEV APIRVYEVATFYSMRNRSKVGKYHLLVCGTTPCM IRGSREIEEALLNHLGVKRNEVTKDGLFSVGEME CMGCCVNAPMITIADYSNGSEGYTYNYIEDVTPK RVVEIVEMLKRGEKPPPGTQNPNRFKCAPEGGH KTLLGEPKPPPCRDLDAC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
883	sp_138210_mznw	-0.884134	-0.922452	-0.963913	Transducin/WD40 repeat-like superfamily protein; FUNCTIONS IN: nucleotide binding; INVOLVED IN: biological_process unknown; LOCATED IN: CUL4 RING ubiquitin ligase complex; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40 repeat, conserved site (InterPro:IPR019775), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat (InterPro:IPR001680), WD40 repeat, subgroup (InterPro:IPR019781); BEST Arabidopsis thaliana protein match is: TBP-associated factor 5 (TAIR:AT5G25150.1); Has 6115 Blast hits to 4690 proteins in 400 species: Archae - 26; Bacteria - 2036; Metazoa - 1593; Fungi - 1149; Plants - 369; Viruses - 2; Other Eukaryotes - 940 (source: NCBI BLink).	MSLISQNAIIKNNKVKKKSKRNNDESEEPSLLEDR EEVLENEGLDGDAALSLKKRKKRNDAEEKRLLVE EELEMQKLGNFLFGSLHSPVEFGNVVEDENENG GLDVDDSALFVMDRSGNKDVSVYGDDEGDQVD ENGDKQRKPAWLDDEEDTVVNISKVNRLRLRK EEDETEISGSAYVSRLRAHHAKLNPGTGWANLD SKVRTYDSDEESEVENGAVVASGYENVDDILRTN EELVVKRGAKLLPRILEYSRLVDANAEDPSNGPIN AVQFHRNHQLLLAAGLDRKLRFFQIDGKRNKMM DSIFIEDCPIRNASFLPDGSQVIVSGRRKFFYSVDI SKGNVDKIGPLTGREEKSLEVFAVSPDSRTIAFLG NEGYILLVSAKTKELIGTLKMNGTARSIAFTDDGQ QLLSSGGDGHVYLWDLRTRTCLHKGVDEGSING TALCTAPVGNLFAAGSSSGIVNVYNKEEFLGGKR KPIKTLEHLRTEASFMKFNHDAQILAICSTMQKSS LKLVHPSFTVFSNWPPASKTLHYPRCLDFSPAG GFMAAGNAAGNVLLYKLHHYQQA	
884	sp_138410_xiut	-0.912282	#N/A	-0.892089	DNAse I-like superfamily protein; CONTAINS InterPro DOMAIN/s: Endonuclease/exonuclease/phosphatase (InterPro:IPR005135); BEST Arabidopsis thaliana protein match is: DNAse I-like superfamily protein (TAIR:AT3G58580.1); Has 1372 Blast hits to 1328 proteins in 220 species: Archae - 0; Bacteria - 20; Metazoa - 540; Fungi - 247; Plants - 315; Viruses - 0; Other Eukaryotes - 250 (source: NCBI BLink).	MLSVVRVHLPSDIPIVGCELSPYVLLRKPDKSVIN DDVPENAPIDGQFLRYKWYRIQSDKKVPVCSVH PTESATLQCLGCVKGKIPVTKSYHCSPKCFSDA WQHHRVLHDRAASAVNENGGEEDELFGRFNSS GSGVLNTGLPPSASSPNLTNGSTPLYPAAVTQR SGGETWFEVGRSKTYTPSTDDIGHVLKFECVVID AETKQTVGHPMTLTTSRVIPAPTPTPRRMISVSG ADVMGHLDPDGRISSAGTFTVVSYNILADAYATS DTYSYCPSWALSWPYRRQNLLREIVTYRADIVCL QEVQFDHFEEFFSPELDKHGYHALYKRKTSEVT HTVDGCATFFRRDRFSHVKKYEVEFNKAAQSLT EAVVAPAQKKSALSRLVKDNVALIAVLEAKFSHQ GADTTPGKRQLLCVANTHVNVHQELKDVKLWQV HTLLKGLEKIAASADIPMLVCGDFNSTPASAPHAL LAMGKVDPLHLDLQVDPLLILRPHTKLAHQLPLVS AYSSFARGIGPALEQQRRRMDPSTNEPLFTNCT RDFIGTHDYIFYTADSLTVESLLELLDEDSLRKDT ALPSPEWSSDHIALLAEFRCKSRIRR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
885	sp_138600_dhjd	#N/A	-0.905815	-0.883394	Pseudouridine synthase family protein; FUNCTIONS IN: pseudouridine synthase activity, RNA binding; INVOLVED IN: pseudouridine synthesis, RNA modification; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pseudouridine synthase, catalytic domain (InterPro:IPR020103); Has 545 Blast hits to 513 proteins in 206 species: Archae - 168; Bacteria - 1; Metazoa - 125; Fungi - 4; Plants - 38; Viruses - 0; Other Eukaryotes - 209 (source: NCBI BLink).	MGASNADEELRLLDDVLRSLPPAAVNDLLTLRLC PQCILRLFGKYGPIYTHYSRISTVLAEVVQHQPSS DDPLLKESANEDGICSVCFGILQFICVDEKQTLVN KGCASDFAALIYELARKDGHQADSLSLEVSIPPIIV ENENAVRLSMKQKYSSEDWFQEMLQSEFISVKD ALKLSITEPLETLLGVKSGSSSFRVRLTYTSSEMP MKTDNGKENNQGSKKRKIGRRSEITDTEVHLKQ CSDSLQEIKSFRLLEKISACSLEVLCSSTPIYIGGR YLKYSRNVSQSRWIIDDERMGEASVEEIVGTSILP LCKGDNYKFHAAGREDIDVRMLGSGRPFLLEIQN PRMLPSELSLKEMETKINSLEGKLVKVKNLKVVD EQGWTLMREGEAEKQKQYAALVWISRQLEDKD LVTISSLKDLKLLQRTPVRVLHRRSPLEREKIIHSM SIEKMSGSTQYFLLHLCTQAGTYIKEFVHGDLGR TYPSLGSILGCRVEILQLDVTDVKMDCFLTE
886	sp_139080_gwju	-0.883388	-0.953319	-0.884508	cytochrome P450, family 72, subfamily A, polypeptide 7 (CYP72A7); FUNCTIONS IN: electron carrier activity, monooxygenase activity, iron ion binding, oxygen binding, heme binding; INVOLVED IN: oxidation reduction; LOCATED IN: endoplasmic reticulum; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: cytochrome P450, family 72, subfamily A, polypeptide 13 (TAIR:AT3G14660.1); Has 33011 Blast hits to 32901 proteins in 1695 species: Archae - 71; Bacteria - 5150; Metazoa - 11421; Fungi - 6601; Plants - 8312; Viruses - 3; Other Eukaryotes - 1453 (source: NCBI BLink).	MGQTSQDDEPCLPYGEVEGQSLKILLXLVLINAG KNYFLWSGPVPMINISKLELIREAFTKTNEFRKTK VNPIFHKLVPGLVSLKGEKWVKHRRMMNPAFRM EKLK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
887	sp_139630_cwdw	-0.933915	-0.89407	-0.906078	DUO POLLEN 3 (DUO3); CONTAINS InterPro DOMAIN/s: Homeodomain-like (InterPro:IPR009057); Has 1836 Blast hits to 1575 proteins in 289 species: Archae - 4; Bacteria - 248; Metazoa - 589; Fungi - 287; Plants - 112; Viruses - 46; Other Eukaryotes - 550 (source: NCBI BLink).	MKTSPLTDEEKARIQEGLKAFKLDWVSVWRTMV PYRDPNLLPRQWRIAIGTQKSYKADSAKKEKHRV YESNRRRSKMAVVDYFQTGSEKENESGEDCVE NDNEAYVHEAFLADWRPTPTGVVQNGIDNQRN QASQSGKVHAQKVSAATEDSHLVHCKSVHPHTS SFPPPIPISETFGESVFRSQRAGKPNLVKLAPELP HVKLPPAVRVISQASLRSSHLGSSGEVGGVNCS ADLVKHMQDKSMISNPNSLSVTNELLKNGYSLTK ERGANLDPQMHPLLFRTIEEGNLPCYPVNSPMRI PTFSFFPPTLLQVPDAGLKSSSSKKETAPPSSSL DFHPLLQRAKEFHANSLDTGCVLNTDLELSRDQF VEDIVYDAGATAAAQTSPSADANDLDLNIHPSSA SKRRKVSSQQGAEINPSVDGSRVVESVKETSSD FSEFAELQPTADLVAKRHTAEIPNPEKSVAVSHK LVSSNDDVGIQPLLEVVMEQEELSDSDEEWEDIE FECEEMADSDGEKESDAVQVVDADAFHKEQVD DAALVTATHAQLLKSLASEQTQTQRKGRTPRKG AGKGKGHTKGSWLSLNSHHKPLSRIKRTISPNRD CSKEISSSKDVSCSTEPKTGISEKDSVSSPSRRT RKHASKSSSSPLEVNSNSSCVN
888	sp_139820_mofj	#N/A	#N/A	-0.875908	Late embryogenesis abundant protein, group 2; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to cadmium ion, response to desiccation, embryo development ending in seed dormancy; LOCATED IN: plasma membrane, membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Water stress and hypersensitive response domain (InterPro:IPR013990), Late embryogenesis abundant protein, group 2 (InterPro:IPR004864); BEST Arabidopsis thaliana protein match is: Late embryogenesis abundant protein (TAIR:AT1G01470.1); Has 294 Blast hits to 286 proteins in 78 species: Archae - 4; Bacteria - 52; Metazoa - 0; Fungi - 0; Plants - 235; Viruses - 0; Other Eukaryotes - 3 (source: NCBI BLink).	MASDDNPELVERAFKGKDSKKEEKEEEKEGFIEK VKDFIHDIGEKIEGAIGFGKPTADVSGIHVPHISLH KLDIVVDVLVKNPNPVPIPLIDINYCIDSEGRKLISG LIPDAGTIHAHGSETVKIPATLIFGDIKETYKDIKPG SIIPYRLKVDLIIDVPVIGRITIPLEKTGKIPIPYKPDI DIEKIKFKRFSFEETAAIIHLKIENKNDFDLGLKDLD YEVWLGHVNIGGADLEKSEILPQNGISNLHLPVTF KPKDCGSALWDIMRGKGTGYSMKGNINVDTPFG EMKLPISKEGGTTRLTKKREAGGSDDDED

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
889	sp_140140_mgwa	-0.859032	-0.862875	#N/A	APS reductase 1 (APR1); FUNCTIONS IN: adenylyl-sulfate reductase activity; INVOLVED IN: response to oxidative stress, sulfate reduction, sulfate assimilation; LOCATED IN: chloroplast, chloroplast stroma, plastid; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Rossmann-like alpha/beta/alpha sandwich fold (InterPro:IPR014729), Thioredoxin-independent 5'-adenylylsulphate reductase (InterPro:IPR004508), Thioredoxin fold (InterPro:IPR012335), Phosphoadenosine phosphosulphate reductase (InterPro:IPR002500), Thioredoxin-like (InterPro:IPR017936), Thioredoxin domain (InterPro:IPR013766), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: APS reductase 3 (TAIR:AT4G21990.1); Has 5999 Blast hits to 5419 proteins in 1658 species: Archae - 169; Bacteria - 2693; Metazoa - 1288; Fungi - 563; Plants - 600; Viruses - 0; Other Eukaryotes - 686 (source: NCBI BLink).	MALAVTSSLSSELKAPSIGVFQDRQYQNLTATISL TTRRAGSAVKPLRAPAESRRSDSVVPLAATTFAA PKTEEGVKEEVTLVEEEENFEDLAKELANASPLEI MDRALEKFGDDIAIAFSGAEDVALIEYAKLTGRPF RVFSLDTGRLNPETYRFFDEVEKHYGIRIEYMFP DSVEVQGLVRNKGLFSFYEDGHQECCRVRKVRP LRRALKGLRAWITGQRKDQSPGTRSEVPVVQVD PVFEGLDGGVGSLVKWNPVANVEGNNVWNFLR TMNVPINSLHAQGYVSIGCEPCTRAVLPGQHERE GRWWWEDAKAKECGLHKGNLKEESSENQNGG SAQANGEVADIFESENLVTLSRAGIENLMKLEDR RDPWIVVLYAPWCQFCQAMEGSYLELADKLAGS GVKVGKFRADGDQKAFAKSELQLGSFPTILLFPK HSSQPIKYPSENRDADSLLAFVQALR
890	sp_141200_kpwp	-0.936025	-0.943838	-0.984699	DnaJ domain ;Myb-like DNA-binding domain; FUNCTIONS IN: heat shock protein binding, DNA binding; INVOLVED IN: protein folding; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), MYB-like (InterPro:IPR017877), Heat shock protein DnaJ, conserved site (InterPro:IPR018253), SANT, DNA-binding (InterPro:IPR001005), Homeodomain-like (InterPro:IPR009057), Myb, DNA-binding (InterPro:IPR014778); BEST Arabidopsis thaliana protein match is: DnaJ domain ;Myb-like DNA-binding domain (TAIR:AT5G06110.2); Has 60375 Blast hits to 43868 proteins in 3410 species: Archae - 228; Bacteria - 14410; Metazoa - 19184; Fungi - 5541; Plants - 3499; Viruses - 192; Other Eukaryotes - 17321 (source: NCBI BLink).	MMAQQRSHVLLISYTQELKDGKPLYVSSNCLPLK VSKFEPAGHSFHSAALRLSGYCEEDSDKDDDKS VSNDKETFTPSFESYSGKSKKKSGGDGKQQQD HYALLGLAHLRYLATEEQIRKSYRDVALKYHPDK QAALLLLEKTEASKQAKKDEIENHFKAIQEAYELLI DPQKRRMYDSTDEFDDEIPTDCAPQDFFKVFGP AFMRNSRWSVNQPAPSLGDDNSSFEEVEKFYN FWYSYKSWREFPHEDEFDLEQAESRDHKRWME RQNAKLSEKARKEEYTRVRSLVDNAYKRDPRIAK RKEFEKAEKKRKKESKFLAKKLEEEEAARVAEEE RLKKEEEDKKAAEAAVQQKKAKEKEKKLVRKER TRLRSLASEVASGSAFDISGDDVETLCMSLDMER LKNLCNKLEGKAKEEQVKLLKDSLSGKVSEENGP TPAPNPTPKPTGAVKSNGNVASSSVLEKKEKEK EKPWVKEEIDLLRKGIQKFPKGTSRRWEVISEYI GTGRSVEEILKATKTVLLQKPDSSKAFDSFLEKR KPAVSISSPLSTRDEVPAVNGPVNGPESVPVPEP VADATSSTTTSTDSPNVSSSTDIPSASSDPDVWS SVQERALLQALKTFPKETAQRWERVAAAVPGKT VNNCKKKLASLKETFRSKKTTA
891	sp_143030_jqaf	-0.85541	-0.906302	#N/A	uncharacterized protein	MSLPEVTITLGRCVLVVRLSRLEEKGDITNLFLTT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
892	sp_143320_njor	0.9112332	0.9244573	0.9377447	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G61450.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MDDTGTILFQISTLKDMLDKVNDEIETNIQITREIE SEIVKIEEIETALDARESELMKLIYLSQYELSGFIAV TVNSRKSVAALTEEINHQRRKKADLLERLNKKREI FSTACIEFQKNIESEECNRVRSMLLEKESLEDEIQ KLDRIDTKFRIPMTAVSDEILEDLHSSNSALAMEIQ ERQAEHERLLKEIEQFKSVLLSTISDVDDPW
893	sp_143380_yvmq	0.9385509	0.9686311	0.9900394	hydroxyproline-rich glycoprotein family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cytosol, nucleus, plasma membrane; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 9 growth stages; BEST Arabidopsis thaliana protein match is: hydroxyproline-rich glycoprotein family protein (TAIR:AT5G57070.1); Has 16953 Blast hits to 10626 proteins in 654 species: Archae - 27; Bacteria - 1228; Metazoa - 7030; Fungi - 2381; Plants - 3856; Viruses - 536; Other Eukaryotes - 1895 (source: NCBI BLink).	MEQKDEDPPPFWMQNTTALRRADRHRHRRNQL ASSLFLNTGVLVVLLLAAAVVFLAIVIPSVLSLARSI FNPSLVKSSWNSLNLILVLFAIVCGFLGRKDDDQ STPVSSPMKQEAVKSVESSSPRQWYEYQDSMR RLRRNISSYPDLRQEGQWLVSENRWRSYDDTH VYNHQFQRSESRISRSRSMGVLDEDEIDVKNIGV DTFVGGGIEVAPPSRTPSPPTPPPPAASHNNVHV RSKQVADEDEDQYVTKNIEMVTKSESIPPSVHQS IPPSPPLSPLSPISPTSYFWPSPSPLPPPPPPPPQ RAHSSMSVKKKKHETKKKHENNPGKRESRSKVP IQAPPPPPPPPPQSRPIRGQFSEDTRDKKRGIV VRTLSRLPRRKRKQARRSYENLTEIIQPPKTSIQR SKSTPEFPPASPPHPPPPPPPPPSMFQNIFTKK GKGKRKVHSHFTTTSSPRHVPSPPARPQQHQY QSHRPSPKTTTTPPATRRTVPLPPAKTDSFSSMA GNNGNGSESPLIPIPPPPPPPPFKMSAWRFVTQ GDYVKIESNNNSPCSFGFSASEDEDSSPNNQKS SNLFCPSPDVDTKADDFIAKFRARLKLDKMNSIKR NHGTGPSPLGSDSGQP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
894	sp_143570_pccr	-0.8810043	-0.8266804	-0.8063656	ubiquitin-protein ligase 1 (UPL1); CONTAINS InterPro DOMAIN/s: Ubiquitin-associated/translation elongation factor EF1B, N-terminal (InterPro:IPR000449), Ubiquitin-associated/translation elongation factor EF1B, N-terminal, eukaryote (InterPro:IPR015940), E3 ubiquitin ligase, domain of unknown function DUF913 (InterPro:IPR010314), Ubiquitin interacting motif (InterPro:IPR003903), E3 ubiquitin ligase, domain of unknown function DUF908 (InterPro:IPR010309), HECT (InterPro:IPR000569), Armadillo-type fold (InterPro:IPR016024), UBA-like (InterPro:IPR009060); BEST Arabidopsis thaliana protein match is: ubiquitin-protein ligase 2 (TAIR:AT1G70320.1).	MIGLGSINSCCLSLAQGWGSKEEGLGLYSCVVAN EKSQEEGLSLFPSDVQDKDEKSQTRLGSTLYFEL HGVSPQRSGESSSDVKSSSSTMIHMPDLHLRKE DDLLLLKQCIEQYNIPIEQRFSLLTRIRYAHAFRSL RSGRLYSRICLLAYIVLVQSNDANDELVSFFANEP EYTNELIKIVRSEEVISGTVRTLAMLALGAQLAAY SSSHDRARMLSGSSINFAGGNRLILLNVLQKAILS LKNSNDPSSLAFIEALLQFYLLHVISTSSSGSIIRG SGMVPTFLPLLEDSDPTHMLVCYAVKTLQKLM DYSNAAVTLLKDSGGVELLAQRLHIEVQRVIGTA GSDGTWMVLGEGSSSSEDHYYGQKRLIRALLKA LGSATYAPTNTSSSQSSHDSTLPSTLALIFSNVG KFGGDIYCSAVTVMSEILHKDPTCFPILHEMGLPD AFLSSVVAGVLPSSKALTCVPSGLGAICLNTKGL EAVKDSSALRFLLEIFTSKKYVVAMNEAIVPLANG VEELLRHVSSLRNTGVDIVIEIINKISSLGEGSSSG SSEKVNESNSMEMDSGEKGSGGESSSSTVTPS DSAGEDITDEHFINLSVFHVMVLIHRTMENSETCR LFVEKSGIDALLKLLLRPSIAQSSEGMAIALHSTM VFKGFTQHHSSPLARAFCSFLRDHLKKALTGFSL VAGSFLLDSKTKLDHGVFSSLFLVEFLLFLAAAKE TRWISALLTEFGGDNKDVLEDIGRVHRETLWQIA LLEDAKVEVQEGAGSVSASESQQVSSSAGPEEQ RVTSFRQLLDPLFRRRMSGWSVESQFFDLINLYR DLGRAGGIQRSTSDSSNLRLGSSQQSRESAGSD TDIALRKEEDKQKIYYSSCCDMVRSLALHIMHLLQ ELGKSMLQPSRRRDDAINISPSSKAVVSTFASICL GHMNFEGHMSPSGSEASVSSKCRYFGKVVDFID GILMDRHDSCNPVLLNCLYGQGVVQSVLTTFEAT SQLLFTVPASPMETDDGNLKHEKKEESEHLWLY GPLASYGKLMDHLVTSSYILSPFTKHLLVQPLANS NVPFFQDAETFVKVLQSMVLKVVLPVWTHPQFV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
895	sp_143960_yemn	-0.8977883	#N/A	#N/A	NAD(P)-binding Rossmann-fold superfamily protein; FUNCTIONS IN: oxidoreductase activity, binding, catalytic activity; INVOLVED IN: oxidation reduction, metabolic process; LOCATED IN: endoplasmic reticulum, plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Short-chain dehydrogenase/reductase, conserved site (InterPro:IPR020904), NAD(P)-binding domain (InterPro:IPR016040), Glucose/ribitol dehydrogenase (InterPro:IPR002347), Short-chain dehydrogenase/reductase SDR (InterPro:IPR002198); BEST Arabidopsis thaliana protein match is: NAD(P)-binding Rossmann-fold superfamily protein (TAIR:AT3G03350.1); Has 112445 Blast hits to 112247 proteins in 3561 species: Archae - 908; Bacteria - 75731; Metazoa - 6658; Fungi - 6086; Plants - 2548; Viruses - 2; Other Eukaryotes - 20514 (source: NCBI BLink).	MLPLPVLFLILVILYIVKFAIVDADFTLRSKKLKRKE FEDKVVWITGASRGIGEVLAKQFASLGAKLILSAR NEAELERVKNQLSGKHAPHDVKILPLDLASGEES LREAVEKAEAFFPGGVDYMIHNAAYERPKSAALD VPEESLKATFEVNVFGTISLTKLLAPYMLKRGRG HFVVMSSAAGKTPAPGQAMYCASKFALNGYFHC LRSELFQKGIKVTVVCPGPIETSTNAEAGSSGQA KSIEKRVSAERCAQLTIIAASHGLKEVWISYQPVL GVMYLVQYMPTIGFWLMDKIGEKRVETAAQKGN TYSLRLLFGKSKRE
896	sp_144160_wuxy	-0.9247227	-0.877376	-0.8506677	translocase inner membrane subunit 17-2 (TIM17-2); FUNCTIONS IN: protein transporter activity, P-P-bond-hydrolysis-driven protein transmembrane transporter activity; INVOLVED IN: protein transport, protein import into mitochondrial inner membrane; LOCATED IN: mitochondrial outer membrane, mitochondrion, mitochondrial inner membrane, mitochondrial inner membrane presequence translocase complex; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Mitochondrial inner membrane translocase complex, subunit Tim17/22 (InterPro:IPR003397); BEST Arabidopsis thaliana protein match is: translocase inner membrane subunit 17-1 (TAIR:AT1G20350.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MGTPETSREPCPDRILDDLGGAFAMGAVGGSVF HYVRGSINAPKGERILRALETVRMSAPRTGGGFA VWGGLFSTFDCTLVYIRQKEDPWNSIFAGAATG GFLQVRQGLRASGRSAAFGGILLALIEGAGIMLN RVMSPPAQIMYEDPSMGAGAPQQMGGYGFGFP GQAANVGGVLESFDTPTPMPNFDYK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
897	sp_144200_quwk	#N/A	#N/A	-0.8838122	SIZ1 ; CONTAINS InterPro DOMAIN/s: DNA-binding SAP (InterPro:IPR003034), Zinc finger, MIZ-type (InterPro:IPR004181), Zinc finger, PHD-type, conserved site (InterPro:IPR019786), Zinc finger, PHD-type (InterPro:IPR001965), Zinc finger, FYVE/PHD-type (InterPro:IPR011011), Zinc finger, PHD-finger (InterPro:IPR019787); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein (TAIR:AT5G41580.1); Has 1186 Blast hits to 1185 proteins in 207 species: Archae - 0; Bacteria - 6; Metazoa - 654; Fungi - 241; Plants - 117; Viruses - 0; Other Eukaryotes - 168 (source: NCBI BLink).	MCESGSNGSNARKQNCSGSSRTQPREEAEESN ELERKLRCPCGSDVPGGGPLVQCLDPDCNVWQ HSSCVIIPEKPLEKISMPIHFLCEICRVKRADPFWK TMAHPLYPCKLGSSALSIDGTNPHMILEKNFKLTK ADRELLQKAEHEIQVWCILFNDRVPYRMQWPLY AELLVNGKPVRTVSRATGQLLGANGRDDCAKIM AYIVEGVNKISLSGTDARSFCFGVKLVRRRTLEQ MSGSRMRTAGRFRACAHMGCFDLQTFVGLNER SRKWQCPICLKNYCLEDVIVDPFFNLILKKMQRC GEEIHEIDMKPDGSWRVKNAREVGDLAQWHFP DGLIQVSSISQTAREHVSERNTSKLTQDNVANW HAVGQGSDEVGNIDQNIITMSSTTTGDLKDDGDP SINQDGVEQFGVSDMNEPEITSGENCNLPSAFAD VIVLSDSEDDDVNLVHPETSHVVVCHTKNEECLISS PGENLNSYPDVASDEGPGIFRGSAGDFLIPQSPT PSRLPEGSGFDFFGTSGPLLDEHNSLTSLTYPRP DNCELTAASAVDPAHHVIPSYDPLNTSFDHQFAS ANGTIPLEAILPCTLISGNELPILNSSPGEDWISLK TGANGNESSSSNVEAHQESADENGLDLQNQFR SKEDSLLQRTRGESGSRSDIAASRKRSDGPFTFP RQQRSQRRRTCVQTSKDLASKQS
898	sp_144590_rgox	#N/A	-0.8890618	-0.8827341	Plant Tudor-like RNA-binding protein; FUNCTIONS IN: RNA binding; INVOLVED IN: biological_process unknown; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 10 growth stages; CONTAINS InterPro DOMAIN/s: ENT (InterPro:IPR005491), Tudor-like, plant (InterPro:IPR014002), Agenet (InterPro:IPR008395); BEST Arabidopsis thaliana protein match is: Plant Tudor-like protein (TAIR:AT2G25590.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MMRFKKGTKVEVLTQKDVPSGSWRCAEIVGGN GHNYSVKYEGDVELSGNGTVERVSRKAIRPCPP SVDVLQDWVSEDVVEVLHNFSWKMAIVLKVLGL NYLLVRLVGSTVELKANKFDVRARLCWQDGEWA VIGKGSNKGEDSRFRIPSSRKGNNQSKKIDAIVKR PTKGYQISTSNTDNALESHVVPSRTLKRKQSHCH FKIEAYAGPRQKTPVEKRDNMPHQLNFVPRQTQ LPDKVDDVASSREMLHEKGMHPSFTNKVTGYSE LIWEREKKDGIVGCSRSISLESDDDNSVSSSVAS CSVYGNYSHDFHRVFATGPYEDNDSQNSDAESF SQLRPVEENSFLPAKEELTEELHRLELHAYRCTIG ALHASGPLSWEQELLLTNLRHSLHISNDEHLKEIR NLVSSSTCIPSS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
899	sp_145200_eggu	-0.883883	-0.883344	#N/A	uncharacterized protein	RHEQGKTRAGGEERKERREEGKKEEGTAAQPG RRDSEAEGKGTAGRGGDGGCRVAVYAEIIMNSS VFADITLKFWHGGRFKRVGVGELVYMGGKCRTV AVDPDELCYFDLVDLAKKCGQYEKIAGLYYRVQG ISLEDGLRQIFNDEEVAEIGELVVKCRTIELYVLHC DDEIVHPDTDPKHSQPLTHPSPFAGNAKKLTPKR APPKSLMSPTKSSPRRQFQTTASNTNDNVTLSS PLKKTQTLASTLSKKDYFSTKYVASSSLVEPSLYK QIENPDVNTQAAQNTNFLSQNTNSTYDSSIPSDY EWEDPRPESPLKYNELISEDSDDDHDPLYKPDS GVLFHDSEDDDLNYEGEEGLNLEFEEGEEEEEE NGRGSRGRGQGDDVGAFGRGSSGRGQGGGCA ATGRGKGKGRGRGKGKNQVPVGVGVYIGADGT PFSNVS
900	sp_145430_gtrw	#N/A	-0.883057	#N/A	isopentenyltransferase 1 (IPT1); FUNCTIONS IN: AMP dimethylallyltransferase activity, ATP/ADP dimethylallyltransferase activity; INVOLVED IN: cytokinin biosynthetic process, secondary growth; LOCATED IN: chloroplast; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: 4 anthesis, C globular stage, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: tRNA isopentenyltransferase (InterPro:IPR002627); BEST Arabidopsis thaliana protein match is: ATP/ADP isopentenyltransferases (TAIR:AT3G19160.1); Has 7944 Blast hits to 7761 proteins in 2606 species: Archae - 0; Bacteria - 5388; Metazoa - 167; Fungi - 152; Plants - 293; Viruses - 0; Other Eukaryotes - 1944 (source: NCBI BLink).	MAPTASGKSRLSVEVASRYPSEIINSDKIQVYKGL DITTNKITLEEQLGVPHHLLGNIDSVIHPELSAAEF RSRASSIISEISGRGNLPIVVGGSNSFVYALLANE FNKDLNVFSESGAQVCSELRYNCCLIWIDVSLPV LDEYIKKRVDDMIDSGGLNELAKFHQYDMLHLEP NVGLTKSIGVPEFKEYVSRYGNEMPSSKGDQKQ KDLYDKAVNLMKENTCKLAKAQMEKIKFLKQRG WDIKRLDATKSFMALLQGSKNWSTIWETQVLSV VGLLQDEVEVSVMKVEKAPLESFADIGGLDPQIQ EIKEAVELPLTHPELYEDIGIKPPKGVILYGEPGTG KTLLAKREAGAIYNRERTGNLTIERGKGGQETVL VKKEKSPKLNPTELRKLLCELGDHLFTGTCYFSP KNHMTKDPRPKVPSA
901	sp_145440_nqwd	-0.882436	#N/A	#N/A	uncharacterized protein	MNSERRTAITKCSHTEKNKNTDEKEEEETSITGT DP

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
902	sp_146110_fhiw	-0.92358	-0.86714	-0.9883046	Vacuolar sorting protein 39; FUNCTIONS IN: small GTPase regulator activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Citron-like (InterPro:IPR001180), Vacuolar sorting protein 39/Transforming growth factor beta receptor-associated domain 1 (InterPro:IPR019452), Vacuolar sorting protein 39/Transforming growth factor beta receptor-associated domain 2 (InterPro:IPR019453); Has 401 Blast hits to 372 proteins in 133 species: Archae - 0; Bacteria - 2; Metazoa - 168; Fungi - 105; Plants - 74; Viruses - 0; Other Eukaryotes - 52 (source: NCBI BLink).	SFKTRDHRDLSPAVQYSSTAVTLRRDSPLPLPEN ISEKFAGKHFSFPGASSMAKKLETSKSRTVLKSL AYFDLSLTSPSSFSLSSLPSFTCVAVATLSDSQLV IYAGTDSGALLLLTLNPDLDPSSSTAAQIDADYNK VKFLRSIPISRSALKSIHIVDQIGRVLVLSDGFLYLV DLQLLLQPVKKLGILKGVNVVARRVCSTEFGSSN WFHEEGVSNSSVGGRLMKKLSVGIRSNGVGVK EFDLRGDNNDGACLFVVTMGKKVVFIELNLSISK HDRDVDGQRGSFVVLKEMPCIDGVKTMVWLDD SVILGTDNAYSLMSCNTGQVSLIFSLPDMAASPC LKLLSKDKVLLLVDNAGITLDSNGQPVGGSLVFG CAPHSIAELSTYVVAGDGKMELYHKRSGTCVQ TFSFGGNFGSSKCYLADEEYVNGRFVVVAASSK VFCYFRKIAPEEQIKDLLRKKNFREAISLVEELEVE GDLSREMLSFVHAQVGFLLLFELHFEKAVDNFLL SETMQPSEIFPFVIRDPNRWSLLIPRNRYWGLHP PPVHIEDVINDGMLAIERAPFLKKAGVDTAVNNEF LLNSPNRDDILDSAYRNIIRYLQVMRNKNINQQVK EGVDTLLMYLYRILKWTEYMEKLASSENSCVIEE LESLLEESGHLRTLAFLYANKGMSSNALAIWRVL ARNHPPGTHQDTAETYHLSDSCSIPVSSRETAAT EASKILEESADVESILQHLGWISEVNEELAVRVLT SQKRISPLPPDDVVAAIDPRKIEILQRYLQWLIED QDSTDTHFHTLYALSLAKTALESLETDVSQHANN KSMREPITYVSGRILPSENPVRERLQIFLHSSDSY DAEEVLELIEGSELWLEKAILYRKLGQETLVLQILA LKLEDSEAAEQYCVELGRPDAYMQLLLDMYLDPQ NAFDFSSITQRLSPDMPLQLASETVLRMLRARLH HYHQGKIVHSLSRAVDVDARLARLEERARNVVLN DDSLCDSCHARLGTKLFAMYPDDSIVCYKCFRRL GESTSVARRNFKKDALIKPGWLVLR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
903	sp_146150_zkdp	-0.851899	#N/A	#N/A	translocase inner membrane subunit 17-2 (TIM17-2); FUNCTIONS IN: protein transporter activity, P-P-bond-hydrolysis-driven protein transmembrane transporter activity; INVOLVED IN: protein transport, protein import into mitochondrial inner membrane; LOCATED IN: mitochondrial outer membrane, mitochondrion, mitochondrial inner membrane, mitochondrial inner membrane presequence translocase complex; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Mitochondrial inner membrane translocase complex, subunit Tim17/22 (InterPro:IPR003397); BEST Arabidopsis thaliana protein match is: translocase inner membrane subunit 17-1 (TAIR:AT1G20350.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MGTPETSREPCPDRILDDLGGAFAMGAIGGSVF HYVRGSINAPKGERILRALETVRMSAPRTGGGFA VWGGLFSTFDCTMVYIRQKEDPWNSIFAGAATG GFLQVRQGLRASGRSAVFGGILLALIEGAGIMLN RVVSPPAQIMYEDPSMGAGAPQMGGYGFGFPG QARAPANVGGVLESFDTPTPMPNFDYK
904	sp_146630_dzru	-0.886481	#N/A	-0.863787	unknown protein; INVOLVED IN: N-terminal protein myristoylation; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G23930.1); Has 4658 Blast hits to 3455 proteins in 431 species: Archae - 39; Bacteria - 387; Metazoa - 1842; Fungi - 436; Plants - 133; Viruses - 38; Other Eukaryotes - 1783 (source: NCBI BLink).	MGVCSSKMKRGRYYRKDDPPMFIQKMKQLQQEI SAILKQREEETEVYERELMLFAFRETEWKRDRKK LREEVKGLRRSLEDREQRIRRMEVEPGITDKSEQ DNHTTHTSTTTTTTTSSSFLLVENMREEERATRDE AVEKWKTLYLAIKHELDHLIQRTHQGTLYWRADD EDLMEDLQKELKAKDETIQVLKAQLASMEDEDYK RKREVDILRQSLRIMSINSSNKGPTSSAKVKKLST NSIRFTKKLAFVN
905	sp_146991z5722	#N/A	-0.866386	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
906	sp_147060_zmpi	-0.8742851	#N/A	-0.8853187	Protein kinase superfamily protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: plasma membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT1G79640.1); Has 122716 Blast hits to 120970 proteins in 3289 species: Archae - 117; Bacteria - 14305; Metazoa - 44836; Fungi - 12007; Plants - 31268; Viruses - 500; Other Eukaryotes - 19683 (source: NCBI BLink).	QLTSPFRLLLPPPSGVLIASSSSALRRSDCLLLYIS CIITSTRRCLCHCRFLIWETFNLSFSAAAIFGVGR CYIFLLLLSLCWFKKALTIGVLHYYLXDAKDYKLY EEVGEGVSATVYRALCIPFNEVVAIKVLDLERCN NDLDGIRREVHTMSLIDHSNVLRAHCSFTAGHSL WVVMPYMAGGSCLHIMKSTYPEGFEEPVIATLLC ETLKALLYLHAHGHIHRDVKAGNILLDTNGTVKLA DFGVSACMFDTGDRQRSRNTFVGTPCWMAPEV MQQLHGYDFKADIWSFGITALELAHGHAPFSKYP PMKVLLMTLQNAPPGLDYERDKRFSKSFKDMVA ACLVKDPKKRPTSEKLFKHPFFKHAKSYEFLARTI LDGLPPLGERFRVLKVKEADLLVQNNALNGDKE QLSQQEYIRGISAWNFNLEDLKSQAALIQDDDSS STAENPDLGLNRMDILDDQSFPTDKPSSEGTNH SNTSNSHEDGFDDLDSLEGSLASFPIQPLQALKG YFDIGEDDMKSIGYKNASLPDAKVHFQKTENEQL GRTDGDNLVRSSSLPQFVSTGQKKVLSGPLLYE NASSSWKQIGDGDRDNMQPRHPSDRNYSGPIS YRQRRDTNNNTLVEDSSEGSVVQRKGRFQVMS DATPKGSTCLSSVPGSTAGLTPPYLSASAILPTLQ YILQQNNLQRDEITKLIKHVELDNLCESADAMSNG ISQTLPISVKEGELLTQVMQLQQRIGSLVDELQR QKLKNAQLEKQLNAFVGRPEKTKD	
907	sp_147390_mncm	-0.8661176	#N/A	-0.8583819	uncharacterized protein	MVARKQFDGVRDRATVVLGVKKEFIAKLYDSRLV DEL
908	sp_147820_pwmt	-0.8704473	#N/A	#N/A	Ribosomal protein L4/L1 family; FUNCTIONS IN: structural constituent of ribosome, rRNA binding; INVOLVED IN: translation; LOCATED IN: ribosome, intracellular, chloroplast, large ribosomal subunit; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Ribosomal protein L4 (InterPro:IPR015498), Ribosomal protein L4/L1e, bacterial-type (InterPro:IPR013005), Ribosomal protein L4/L1e (InterPro:IPR002136); BEST Arabidopsis thaliana protein match is: ribosomal protein L4 (TAIR:AT1G07320.4); Has 8236 Blast hits to 8236 proteins in 2694 species: Archae - 70; Bacteria - 5573; Metazoa - 121; Fungi - 129; Plants - 100; Viruses - 0; Other Eukaryotes - 2243 (source: NCBI BLink).	MAKRILASLHHHLLRPHAHLPSLGVSCSRLISTTT VTSNSSNDEIPSDLLMPKKTAFIPDRNQGVYQDV VIPVTNFHFENKGLMSLAGDVFDVPIRKDIVHRVV RWQLAKRQQGTHSTKTISEVSGTGRKPHPQKG TGRARQGSRRSPHMRGGSTMHGPKPRSHAFKL NKKVRRLGLKIALSARAAEGKLVIFDDLEVPTHKT KNIVHWVNGLDDSKKLLLVDDGVINEKLKLATQNI HYVNVLPSIGLNVYSILQHDTLVMSRAAVNRIVER MNTPIKR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
909	sp_147880_qyhp	-0.968383	-0.972512	-0.978486	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: protein kinase activity, kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Leucine-rich repeat protein kinase family protein (TAIR:AT2G37050.3); Has 172111 Blast hits to 130582 proteins in 4641 species: Archae - 151; Bacteria - 14946; Metazoa - 47791; Fungi - 11011; Plants - 76783; Viruses - 518; Other Eukaryotes - 20911 (source: NCBI BLink).	MGSYFHVLALLLLSLPCIFSQVTEFISIDCGGDSN YTDKKTGLAWISDTGIMSHGETIDVQNTNDEWE PYKKRRDFPTDDKKYCYNLETKERRRYLVRATFL YGSTENADTYPSFLLYLDATKWATVTVFDGSRTY VEEMIIRAPSSSINVCICCATTGSPFISTLELRPLNL SMYATDYEDHFYMKSAARVNFGAPTKDPIRYPD DPFDRVWESDLVKRPNFLVGIAPGTERINTTKNIE VTNREYPPVKVMQTAVVGTEGNLTYRLTLEDFP GNARAFAYFAEIENIRSNETRKFKMEHPYVPDYS NAVVNIEENANGTNKLYEPSYMNVTLDFVLSFSF VKTPDSTKGPLLSAIEISKYVQIGSKADAQDVIALN AVLAMLPAGVEKQDEGDPCTPTQWEWVTCSST TPPRITAIHLSKRNLTGPIPPEIQNLDAVTELWLEG NSLIGPIPDISNLVNLRIVHLENNRLNGSLPSYLSS LPSLQELYIQNNDFSGKISPALLSRKMVFNYDGN PKLTKGIRLGTSLKVGISVAALGALIVVILVILLLLR NFKRKTPRQKINYKADANTKPPKKGHNLTRGAHL MEDGIACCIKLSELDAATMKFSKNIGRGSFGVVY YGRMKDGKEIAVKIIGDSTSQGTKQFLTEVSLQV DWLTRLRIAEDAAKGLEYLHTGCSPSIIHRDVKTS NILLDINMRAKVADFGLSRQAEEDLTHVSSVAKG TLGYLDPEYYANQQLTEKSDVYSFGIVLLELISGR KPITEDYGADWNIVHWARSLIRKGDVGSMMDPL LADNVKLESVWRVAEVAIQCVEQHGISRPKMQEI ILAIQDAVMLEKGSENNQNMSSGSGRTPSSRKTL LTSFLEIESPDLSNARLTPSAR
910	sp_148251z5778	-0.86397	#N/A	#N/A	uncharacterized protein	#N/A
911	sp_148610_dwkx	#N/A	#N/A	-0.894899	ORMDL family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: protein folding; LOCATED IN: integral to membrane, endoplasmic reticulum; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: ORMDL (InterPro:IPR007203); BEST Arabidopsis thaliana protein match is: ORMDL family protein (TAIR:AT1G01230.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEKMMYVRAEASKEVNKNTEWVTYPGVWTTYIF ILFFSWLLVLSLLGCSAGMAWTIVHLSHFFVTYH CFHWKKGTPFSDDQGIYNGLTWWEQIENGQQL TRNRKFLTVIPVVLYLIASHTTDYQNPMLMLNTLA VFVLVVAKFPRMHRVRIFGINGEN
912	sp_148750_yecd	#N/A	-0.905726	#N/A	uncharacterized protein	VYKESRKRVPGRTYLTNPEKIEENIEKIDALESTQ DGEGGSNSEDLISEVIQGPKRKRKSRVPLYGKG VTKSDLKKMGKKIRILDSVRVLAEYENRVGASTC TPRCVYDCISTSRS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
913	sp_148970_rkyg	-0.887244	-0.8807765	-0.9455364	SU(VAR)3-9 homolog 4 (SUVH4); FUNCTIONS IN: double-stranded methylated DNA binding, methyl-CpNpG binding, methyl-CpG binding, histone methyltransferase activity (H3-K9 specific), methyl-CpNpN binding; INVOLVED IN: maintenance of DNA methylation, histone methylation, peptidyl-lysine methylation, histone H3-K9 methylation; LOCATED IN: nucleus; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: SET domain (InterPro:IPR001214), SRA-YDG (InterPro:IPR003105), Pre-SET domain (InterPro:IPR007728), Post-SET domain (InterPro:IPR003616); BEST Arabidopsis thaliana protein match is: SU(VAR)3-9 homolog 6 (TAIR:AT2G22740.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEVVVEVGDKVVAAEGRRTSDRTRGRRISYNETI LAKIEEVSPKKRKYQRNGGKRKTVKVPEQKESTE MEMVRGLLGVDGDGLTGLGFKLEEVEEDGNGV VEKSDHLRVKETLRIFNNYYLHLIQDEENRCKLIE AEKKQKAQKGVEDKRIPHKRPDLKAITKMMNTG KVLYPSKRFGAIPGVDVGHQFFSRAEMVAVGFH SHWLNGIDYMGTSYSTGEYSEYTFPLAVAIVLSG CYEDDLDNSEDVVYTGQGGNNLLGDKKQNKDQ KMLRGNLALKNSMEQDLPVRVVRGHSSKSKSYVG KVYTYDGLYKVVNYWAEKGISGFTVYKFRLKRLE GQPELTTNQVQFIRGRIPDSVSEVRGLVCADISE GQENIPIPATNLVDDPPVPPTGFKYRKSIQVDNV KIPAASSGCDCVGGCTNPNICACARLNGSDFPYV HRDGGRLIEAKAVVFECGPKCGCGSSCINRTSQ KGLRYRFEVFRTPKKKGWGVRSWDYIPAGAPICE YTGFLKRSDEGDSVLENNFIFDIDCLQTMKGLDG RQMRLRDVSIPSDGIHEKPEEKVVDSMVEYCIDA GVTGNLARFINHSCEPNLFVQCVLSAHHDIKLAR VMLFAADNIPPLQELTYDYGYALDSVIGPDGKVK EMACYCGAEGCRKRLF	MEMVRGLLGVDGDGLTGLGFKLEEVEEDGNGV VEKSDHLRVKETLRIFNNYYLHLIQDEENRCKLIE AEKKQKAQKGVEDKRIPHKRPDLKAITKMMNTG KVLYPSKRFGAIPGVDVGHQFFSRAEMVAVGFH SHWLNGIDYMGTSYSTGEYSEYTFPLAVAIVLSG CYEDDLDNSEDVVYTGQGGNNLLGDKKQNKDQ KMLRGNLALKNSMEQDLPVRVVRGHSSKSKSYVG KVYTYDGLYKVVNYWAEKGISGFTVYKFRLKRLE GQPELTTNQVQFIRGRIPDSVSEVRGLVCADISE GQENIPIPATNLVDDPPVPPTGFKYRKSIQVDNV KIPAASSGCDCVGGCTNPNICACARLNGSDFPYV HRDGGRLIEAKAVVFECGPKCGCGSSCINRTSQ KGLRYRFEVFRTPKKKGWGVRSWDYIPAGAPICE YTGFLKRSDEGDSVLENNFIFDIDCLQTMKGLDG RQMRLRDVSIPSDGIHEKPEEKVVDSMVEYCIDA GVTGNLARFINHSCEPNLFVQCVLSAHHDIKLAR VMLFAADNIPPLQELTYDYGYALDSVIGPDGKVK EMACYCGAEGCRKRLF
914	sp_149110_oyqt	#N/A	#N/A	-0.8538679	SALT OVERLY SENSITIVE 3 (SOS3); CONTAINS InterPro DOMAIN/s: Recoverin (InterPro:IPR001125), EF-HAND 2 (InterPro:IPR018249), Calcineurin B protein (InterPro:IPR015757), EF-hand-like domain (InterPro:IPR011992), Calcium-binding EF-hand (InterPro:IPR002048); BEST Arabidopsis thaliana protein match is: calcineurin B-like protein 8 (TAIR:AT1G64480.1).	MGCSISKKSKANYWDFASLASETMFSLDEVEAV YVLYKKVSCCIVDDGLIHKEEFRQALLRNKGKPDL FVDRVFDLFDLKRNGVLDFGEFVQSLSIFHPKSP AADKILFTFNLYDLQHTGCIEREEVKEMVLAILKE SDLELPDEVIEEIVENAFNEADSNQDGKIDLEEW KEFVGKNPSLLKNMTIPYLMKERNRFKDSAVLAA QTHCAFHLYDVWDTGYIEPDEVRVMILAFLDESD LLLSDDIQTIIDKTFEEADFKGDGKIDLEEWEALV ARNPSILKNMTIPYLKDIPTGFPSFRHRNDTESES KMY
915	sp_149150_iaof	#N/A	-0.8544233	#N/A	GRF zinc finger / Zinc knuckle protein; FUNCTIONS IN: zinc ion binding, nucleic acid binding; INVOLVED IN: biological_process unknown; CONTAINS InterPro DOMAIN/s: Zinc finger, CCHC-type (InterPro:IPR001878), Zinc finger, GRF-type (InterPro:IPR010666); BEST Arabidopsis thaliana protein match is: GRF zinc finger / Zinc knuckle protein (TAIR:AT5G13920.1); Has 109 Blast hits to 86 proteins in 31 species: Archae - 2; Bacteria - 6; Metazoa - 13; Fungi - 4; Plants - 79; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MELTPRSAGSSSSSNSRASLIEIDVKCNHNITAVLR TVKKGSRNNIGKRFYGCPLWPRGDCGFFQWEC SGCAELQSKLSMKEKELTIVEEMVEKMKAKKVG HEEIVRELKAELSHTRIELMRSARVEKYAAMAVLL SWVIFVVILMLIN	MELTPRSAGSSSSSNSRASLIEIDVKCNHNITAVLR TVKKGSRNNIGKRFYGCPLWPRGDCGFFQWEC SGCAELQSKLSMKEKELTIVEEMVEKMKAKKVG HEEIVRELKAELSHTRIELMRSARVEKYAAMAVLL SWVIFVVILMLIN
916	sp_149930_upms	#N/A	-0.9025036	#N/A	uncharacterized protein	MVVAESVPAADAAVAADDVVPAVTVLLRCCSCC KGCCCWRICSLLLLSFFYFLFLHGYLFLSYLG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
917	sp_150040_honf	#N/A	-0.9831969	#N/A	alpha/beta-Hydrolases superfamily protein; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT1G23330.1).	MKRLWLSSLQLVELFVSSWVHLLYGFYIFSTAVA YDLNQALCDYFKPSNINQIEAAAPPQDSSSVDDL PPIVLVHGIFGFGQQRLGGLSYFAGAEKKDERVL VPDLGSLTSIYDRARELFYYLKGGQVDYGEEHSK ACGHSQFGRIYEQGHYPEWDEDHPIHFVGHSAG AQVVRVLQQMLADKAFEGHDTSPNWVLSITSLS GAFNGTTRTYYDGMQPEDGVTLKPVCLLQLCRL GVIIYDWFDIPWLKSYYNFGFDHFNLSWRKIGVW GLVDCLLGNAGPFASGDWILPDLTIQGSMRLNSH LQTFPDTYYFSYATKRTRQLLGFTAPSGIMGIHPL LFIRVLQMCMWRLPADVPPPYKGYRDEDWWDN DGALNTTSMTYPRFPVEHPHHLVTNDKDCRPLQ PGIWYYKIVEGDHILFIVNRERAGVQFDLIYDSIFE RCRKHVFRKRPQTLPNETPQTVTS
918	sp_150990_rzcx	-0.9844886	#N/A	-0.8393837	PETER PAN-like protein (PPAN); CONTAINS InterPro DOMAIN/s: Brix domain (InterPro:IPR007109); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MGRFRHRKRKIVVKPNVKNQQLPTYDQITGRKIP QSFVFSRTKLPPVLRQLEMDLRKLMLPYTALKLK EKKRNSLKDFLNVAGPMGVTHFLILSKTNVGPYL RVAKTPQGPTLTFKIHEYSLAVDVARSQLRPRCP QDLFKNSPLLVLSGFGTGEEHLTLITRMFQHMFP SIDVNTVKJALCQRIVLVHYNKESKRIDFRHYSIRL QPVGVSRRLRKFVQNHQVPDLRNLQDVSDFVTR AGYGSESEADDEAATVSLVSDLGRVNRASTKSA VKLQEIGPRMTLELIKIEDGMCSGQMLFSNDGKC GGKRTQEDENDEEEEEEDSQEEEEEEDDEKEEG DEEE
919	sp_151700_phuf	-0.965115	-0.976323	-0.9335059	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT5G55840.1); Has 45179 Blast hits to 13990 proteins in 287 species: Archae - 4; Bacteria - 20; Metazoa - 378; Fungi - 562; Plants - 43035; Viruses - 0; Other Eukaryotes - 1180 (source: NCBI BLink).	MNRLSSSFAAPRILLSNYSTKSALTPSSTSFNPPI VPNSKPSLNPLYSFLPETQNPNNLVNLICLALKKT NFQITHLANQLQDLIPHLGTQEISRVLLRCQCESQ SALIFFNWVKNDLGVKPNCHNYCIIVHILVWSRNF VHGMKLLSELIRFQSDESEKVDIFQGLVSCSEDC NWDPVVFDMLIKAYMKDDLIKEGYRVFKKMVKM GYVPHVLAFNCLLSGLSKANCVDNCWELYKEM WRIGINPNTCTFNIFTNVLCKDQSVDKVNEFLEK MEEEGFVPDLVTYNTLIGNYCRSRRLDDAFYLYKI MYRRGVVPDLVTYTSLMNGLCKIGKLKEAHQVF HRMVDRGLTPDIVSYNSLIHGYCIEGRMKDSRLL LHDMIAKGISPDTFTCQVLVEGFVKAGSLRSCLN MIVELKRFGVTISSDVYKYLVIALSQENHPIAAVKL LDRMLEDGHESDVEIYDKLIKSFCDSELLEGALLH KSKMLERRVKPCLAIYQNLISCLCRLRRCLEAEVL MQEMDLFDLKPDNNICRALIFAYCTENNFDKAES LLLSLSIKYQLYDTDSYNVLLKIRSEQSSITELMEL QDKMLKLGFAPNGLTFKYYLQGLQNTVMKDKGS FQACSQLQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
920	sp_152040_tcyr	-0.946184	-0.915211	-0.881137	LOCATED IN: cytosol, plasma membrane; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Beta tubulin, autoregulation binding site (InterPro:IPR013838), Misato Segment II, myosin-like (InterPro:IPR019605), Tubulin/FtsZ, N-terminal (InterPro:IPR019746); Has 345 Blast hits to 341 proteins in 161 species: Archae - 0; Bacteria - 0; Metazoa - 131; Fungi - 140; Plants - 55; Viruses - 0; Other Eukaryotes - 19 (source: NCBI BLink).	MREVVTIQVGGFANFIGSHFWNFQDELLGLASD HDMDPVFQNPGVNMDVLYRSGETQEGVLTYTP RLLSVNYQGSLGSMSSHGTLYNQTPELPSPVAT WKGSISTQRSESLNRNLFLQSLYKEEQKESVGK DSKHADSPDNIGDQEIMESLENGTQFWTDFSKV HYHPHSLYEINGLWMDVEGFDNYGAGKDVFAES SRGEEMTERLRFFVEECDHVQGIQCVVDDSGGF SSVAADFLQNIADEYTNTPVLLYTARDPGSFMSS KSRKRNISRNLNDAVSFAKLSPFCKLIVPVGLPFL GGSKASTYLRVNDEKPYHCSAVYAAALHALSLPF RIDSLGPTATSDCTLGALDFNSIVQMLSGQSRQN MVAILDTTMPAPTIGEQLLSKLQTLTPEISDDMED LHSVETLVIHGALKSDEQRASVFEVQDAISAAYE QASSRPRFCHLSATTCPLPIPLLPFPSIFSQTVGRC GQLLSNPSSSSASKGSLDVHSIPMAARLRSSSAV LPFLTNRLENLRKFGIQHGAMGGELLKSWGFGK EELEEMGETLSNMVRTLAPYSDLTSDSD
921	sp_152600_coqp	#N/A	-0.881411	#N/A	LOB domain-containing protein 21 (LBD21); CONTAINS InterPro DOMAIN/s: Lateral organ boundaries, LOB (InterPro:IPR004883); BEST Arabidopsis thaliana protein match is: LOB domain-containing protein 25 (TAIR:AT3G27650.1); Has 1016 Blast hits to 1011 proteins in 25 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1016; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MKKCYNNDLRSTSSCAACKLLKRRCIPNCIFAPH FRADEPKKFAKVHKVFGASNVSKILNEVPEEHRE DTVNSLSYEAEARLRDPVYGCIGAVVLLQRRMFQ LQHDLALARARLAHYKASSLHTNTNNVVTANINT DDNGLFGLDDHNTVSITNMINLCGEELISSDQLSQ SDIGVYNTTDFGQVPDLLPL
922	sp_153960_hwei	-0.933328	-0.967496	-0.957851	thylakoid lumen 15.0 kDa protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: thylakoid, thylakoid lumen, chloroplast thylakoid lumen, chloroplast; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVSHVLQGEVSKDFFEGEDASIEAAVNAISYFLR EPVSLSNCSEVM

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
923	sp_153970_gmxy	-0.93867	-0.97417	-0.98737	eukaryotic elongation factor 5A-1 (ELF5A-1); FUNCTIONS IN: ribosome binding, RNA binding, translation elongation factor activity, translation initiation factor activity; INVOLVED IN: translational initiation, xylem development; LOCATED IN: cellular_component unknown; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold (InterPro:IPR012340), Translation elongation factor, IF5A, hypusine site (InterPro:IPR019769), Translation protein SH3-like, subgroup (InterPro:IPR014722), Translation elongation factor, IF5A (InterPro:IPR001884), Translation elongation factor, IF5A C-terminal (InterPro:IPR020189), Translation protein SH3-like (InterPro:IPR008991), Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), KOW (InterPro:IPR005824); BEST Arabidopsis thaliana protein match is: eukaryotic elongation factor 5A-3 (TAIR:AT1G69410.1); Has 1356 Blast hits to 1355 proteins in 415 species: Archae - 264; Bacteria - 0; Metazoa - 366; Fungi - 246; Plants - 261; Viruses - 0; Other Eukaryotes - 219 (source: NCBI BLink).	MSDEEHQFESKADAGASKTYPQQAGTIRKNGYI VIKNRPCKVVEVSTSKTGKHGHAKCNFVGIDIFN GKKLEDIVPSSHNCDVPHVNRTDYQLIDIAEDGFV SLLTESGGTKDDLKLPTDETLLKQLKEGFADGKD LIVSVQNAMGEEQICGVKDIGGGSK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
924	sp_154380_osza	-0.887576	-0.949228	-0.872845	NADH-dependent glutamate synthase 1 (GLT1); CONTAINS InterPro DOMAIN/s: FAD-dependent pyridine nucleotide-disulphide oxidoreductase (InterPro:IPR013027), Glutamine amidotransferase, class-II (InterPro:IPR000583), Aldolase-type TIM barrel (InterPro:IPR013785), Glutamate synthase, alpha subunit, C-terminal (InterPro:IPR002489), Glutamate synthase, NADH/NADPH, small subunit 1 (InterPro:IPR006005), Adrenodoxin reductase (InterPro:IPR000759), Fumarate reductase, C-terminal (InterPro:IPR012285), Glutamate synthase, central-N (InterPro:IPR006982), NAD(P)-binding domain (InterPro:IPR016040), Glutamate synthase, eukaryotic (InterPro:IPR012220), Alpha-helical ferredoxin (InterPro:IPR009051), Glutamate synthase, central-C (InterPro:IPR002932), Glutamine amidotransferase, type II (InterPro:IPR017932); BEST Arabidopsis thaliana protein match is: glutamate synthase 1 (TAIR:AT5G04140.1).	MSVAAGVGSVLNLRSKVGSDSFTSNKQTQNHG GLAVVPPQLCSFGKRSSRSYAVRRSEAKICESKF FGNRLRSERLRPLVRDPNARFVIQSALSQVPEQP LGLYDPKYDKDSCGVGFVAELSGENSRKTVTDA VEMLVRMAHRGACGCETNTGDGAGILVGLPHDF FTEVTKDLGFELPALGKYAVGMFFLPTSESRREE SKNVFTKVAESLGHTVLGWRSVPTDNSGLGKSA VQTEPVVEQVFLTPTPRSKADLEQQMYILRRVSM VAIRAALNLQHGGVKDFYICSLSSRTIVYKGQLKP SQLKEYYYWDLGNERFTSYMALVHSRFSTNTFP SWDRAQPMRVLGHNGEINTLRGNVNWMRARE GLLKCKELGLSKNEMKKLLPIVDASSSDSGAFDG VLELLVRAGRTIPEAVMMMIPEAWQNDKNMDPK RKGFYEYLSALMEPWDGPALITFTDGRYLGATLD RNGLRPGRFYITHSGRVIMASEVGVVDIAPEDVS KKGRLNPGMMLLVDFDKHIVVDDEALKKQYSLAR PYGEWLKRQKLVLKDIVDSVAEAERVAPAISGAM PVSNQDEDMQNMGIHGLLAPLKAFGYSVEALEM LLIPMAKDATEALGSMGNDTPLAVMSNREKLTFQ YFKQLFAQVTNPPIDPIREKIVTSMECMIGPEGDL TEMTEEQCHRLSLKGPLLSIEEMEAVKKMDYRG WRSKVIDITYPKTLGRKGLEQTLDRICAEARDAIK EGYTVLVLSDRAFSPSRVAVSSLLAVGAVHQHLV KTLERTRVGLIIESAEPREVHHFCTLVGFGVDAIC PYLAVETIWRLQVDGKIPPRSTGEFRSKEELVKR YFKASNYGMMKVLAKMGISTLASYKGAQIFEAVG LSSEVIERCFTGTASRVEGATFEMLASDALKLHE MAFPRRALPVGSAESLALPNPGDYHWRKDGELH LNDPLAMAKLQEAAKSNSVGAYKEYSKRINELNK SSNLRGLLKFKEVNEKVPLEEVEPVSEIVKRFCT GAMSYGSISLEAHSALAIAMNKLGGKSNTGEGG ENPSRLEPLSDGSRNPKRSAIKQVASGRFGVSS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MATTPSSLRFSSASGSYGNLDAINRVLSDCCRGT
						PKEGASSALRRHIEEAAHDLNGEAFSRFMDQLY
						DRISTLLESGEVAENLGALRAINELIDLPLGESASR
						VSRFSNYIRTVFELKRERELVLASRVLGHLARAG
						GAMTADEVEHQIRNALDWLRGERFEYRLLAAVLI
						LKEMAENASTVFNVHVQEFVDAIWVALRDPSLD
						VRERAVEALRACLRVIEKRETRWRVQWYYCMFE
						ATQEGLGKNAPVHSIHGSLLAVGELLRNTGEFM
						MSRYKEVAEIVLKYLEHRDRLVRLSITSLLPRIAYF
					target of rapamycin (TOR); FUNCTIONS IN: protein binding, 1-phosphatidylinositol-3-kinase activity; INVOLVED IN: embryo development ending in seed dormancy; LOCATED IN: chloroplast; EXPRESSED IN: 28 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Phosphatidylinositol 3-/4-kinase, catalytic (InterPro:IPR000403), Armadillo-like helical (InterPro:IPR011989), FKBP12-rapamycin-associated protein, FKBP12-rapamycin-binding (InterPro:IPR009076), Protein kinase-like domain (InterPro:IPR011009), PIK-related kinase, FAT (InterPro:IPR003151), PIK-related kinase (InterPro:IPR014009), Armadillo-type fold (InterPro:IPR016024), PIK-related kinase, FATC (InterPro:IPR003152), Phosphatidylinositol 3/4-kinase, conserved site (InterPro:IPR018936); BEST Arabidopsis thaliana protein match is: Ataxia telangiectasia-mutated and RAD3-related (TAIR:AT5G40820.1); Has 4645 Blast hits to 4114 proteins in 299 species: Archae - 0; Bacteria - 46; Metazoa - 1929; Fungi - 1123; Plants - 507; Viruses - 4; Other Eukaryotes - 1036 (source: NCBI BLink).	DRLREEIVEKLLIAAVADADVSVRLSIFQSLHEDGG
						FDEFLAQADSLSAIFAALNDEDFQVREYAISVAGR
						QLVEALDQIASIPSLLSTIQDRLLECISLVLSRSHH
						AQVRPAAAIGRGNVANTPLLVSDVSGPVLVQLAL
						QTLAHFNFKGHDLLEFGRESVTSYLEDEDAVTRK
						DAALCCCRLVANSISGTSSQFSSARSNRTGGRR
						RRLIEEIVEKLLIAAVADADVSVRLSIFQSLHEDGG
						FDEFLAQADSLSAIFAALNDEDFQVREYAISVAGR
						LSEKNPAYVLPALRRYLIQLLTYLEQSVDSKCREE
						SAKSLGCLIRSCERLILPYIAPIHKALVAKLREGST
						LSANSGIITGVLITVGDLARVGGFAMRQYIPELMP
						LIVDALLDGAAVTKREVAVATLGQVVQSTGYVITP
						YNDYPQLLGLLLKLLNGDLVWSTRREVLKVLGIM
						GALDPYVHKRNQPNLPGSHAEVTHTTTETGQHI
						RSMDDLPMDLWPSFTTSSEDYYSTVAINSLMRV
						LRDPSLSSYHQKVVGSLMFIFKSMGLGCVPYLQK
						VLPDLFQIVRVCDDNLKDFITWKFGTLVSIVRQHI
						RKYLGDLLALVAELWSSFSLPAVNHSVRGLPVLH
						LVEQLCVALNDEFRTYLPAILPCCIQVLSDAEKCN
						DYTYVPDILHTLEVFGGTLDEHMHLLLPALIRLFK
						VDASVEIRRASIKTLTRLIPRVQVTGHISSLVHHLK
925	sp_154470_uali	-0.8837395	-0.843664	-0.88033		LVLDGKNDELRKDAVDALCCLAHALGEDFSIFIPS
						MESSKGSYAKEWYTWEKQLRETLLSNSEYLNSI
					RNAligase (RNL); CONTAINS InterPro DOMAIN/s: tRNA ligase, phosphodiesterase, fungi (InterPro:IPR015965); Has 209 Blast hits to 196 proteins in 23 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 3; Plants - 200; Viruses - 0; Other Eukaryotes - 6 (source: NCBI BLink).	QIPFESAVNSVVEQLNAIVKGDCMTPSSEKRRVG
						AIIYAGISLPVTEINRHLNDIAVSNSMVERFFSDKN
						LAVTKKVHVTLAHKRAHGVIAVANFGQYINRKVD
						VELVALLFSANVAALEVCLGSVDDEKIESKNEWP
926	sp_154920_okfa	#N/A	#N/A	-0.8857119		HVTIWTASGVAAKQASTLPQLVAQGKASRIEIDP
						PVVIQGVVQFF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
927	sp_155680_xsno	-0.873096	-0.940875	-0.95187	SERRATE (SE); FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: in 7 processes; LOCATED IN: cytosol, nuclear speck, nucleolus, nucleus, chloroplast; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, C2H2-like (InterPro:IPR015880), Arsenite-resistance protein 2 (InterPro:IPR007042), Protein of unknown function DUF3546 (InterPro:IPR021933), Zinc finger, C2H2-type (InterPro:IPR007087); Has 5797 Blast hits to 4596 proteins in 474 species: Archae - 2; Bacteria - 583; Metazoa - 2822; Fungi - 981; Plants - 712; Viruses - 137; Other Eukaryotes - 560 (source: NCBI BLink).	MAEVLNMPVEPIDNRRRNKDKPDEPPTDSSPPP PPPPPNRRRDRDSRERRDDRDFDRPPPNRYYD RNNRGGGGGGGERDRDYNNNNNNNNKRNRGL SPSPPHPPPPYRDNRRNNYSPPPRRSPPYPPFK RSRRDEGGYDGRRGSPRGGGYGPGDRRFGYD HPGGYDREMGGRPGYADDRPHGRFRSGGYPG GPSDWDSGRGGYGDSSNPGNAQREGLMSYKQ FIQELEDDVLPAEAERRYQEYKSEYISTQKRTYFE SHKDEDWLKNKYHPANLLTVIERRNELARKVAKD FLLDLHSGSLDMGPSANASTMNKSGQTSEPNSE DETDVNGKRRRHGRAQAKEADISAAPKAHLVSS EPRRIQIDIDLAQALVTKLDSEKRIEDNILSKDDTD KMSRDRSHSGSIGPVIIIRGLTSVKGLEGIELLDTL ATYLWRVHGVDYYGMIETAEPKGLRHVRAEGKN SDVSTNAAEWEKKLDSHWQDRLRSLDPLEVMT AKEKIDAAAVEALDPFVRKIRDEKYGWKYGCGAK SCTKLFHAAEYVHKHLKLKHPELSIDATTKVREEI YFQNYMNDPNAPGGQPVMQQFVPKDKLQRRRP GADNRLKDDRGNRRERDRNGGERFERPENPQA DFPPNKDGPDKGNSDEPIFDNFGGQGMHVAPFA TDIAPPPVLMPVPGAGPLGPFVPAPPEVAMRML REQGGPSPFEGPGRNGMGAPPIILPPSFRQDPR RLRSYQDLDAPEDEVTVIDYRSL
928	sp_156090_qzmp	#N/A	#N/A	-0.900687	regulatory particle triple-A ATPase 5A (RPT5A); CONTAINS InterPro DOMAIN/s: ATPase, AAA+ type, core (InterPro:IPR003593), ATPase, AAA-type, core (InterPro:IPR003959), ATPase, AAA-type, conserved site (InterPro:IPR003960), 26S proteasome subunit P45 (InterPro:IPR005937); BEST Arabidopsis thaliana protein match is: 26S proteasome AAA-ATPase subunit RPT5B (TAIR:AT1G09100.1); Has 33265 Blast hits to 30927 proteins in 3145 species: Archae - 1437; Bacteria - 12044; Metazoa - 4947; Fungi - 3672; Plants - 3322; Viruses - 56; Other Eukaryotes - 7787 (source: NCBI BLink).	MASASATAMVEDPNFQDDQLANLTTEDIARASRI LDNEIRILKEDVQRTTLELDSFKEKIKENQEKIKLN KQLPYLVGNIVEILEMNPEDEAEEDGANIDLDSQR KGKCVVLKTSTRQTIFLPVVGLVDPDKLKPGDLV GVNKDSYLILDTLPSEYDSRVKAMEVDEKPTEDY SDIGGLEKQIQELVEAIVLPMTHKERFQTIGIRPPK GVLLYGPPGTGKTLMARACAAQTNATFLKLAGP QLVQMFIGDGAKLVRDAFQLAKEKAPCIIFIDEIDA IGTKRFDSEVSGDREVQRTMLELLNQLDGFSSD ERIKVIAATNRADILDPALMRSGRLDRKIEFPHPT EEARARILQIHSRKMNVHPDVNFEELARSTDDFN GAQLKAVCVEAGMLALRRDATEVNHEDFNEGIIQ VQAKKKASLNYYA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
929	sp_156270_kwyn	#N/A	-0.834647	-0.8044832	Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain; FUNCTIONS IN: protein serine/threonine/tyrosine kinase activity, protein kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Octicosapeptide/Phox/Bem1p (InterPro:IPR000270), Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine-threonine/tyrosine-protein kinase (InterPro:IPR001245), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain (TAIR:AT1G16270.2); Has 119780 Blast hits to 118332 proteins in 4796 species: Archae - 124; Bacteria - 12838; Metazoa - 46034; Fungi - 10270; Plants - 32306; Viruses - 451; Other Eukaryotes - 17757 (source: NCBI BLink).	MDSNSIITSNNNNNNNMGNLGNDEMVSPRVKFL
930	sp_156450_izyj	-0.8066616	-0.870336	#N/A	Transcription factor IIIC, subunit 5; CONTAINS InterPro DOMAIN/s: Transcription factor IIIC, subunit 5 (InterPro:IPR019136); BEST Arabidopsis thaliana protein match is: Transcription factor IIIC, subunit 5 (TAIR:AT3G49410.1); Has 680 Blast hits to 642 proteins in 201 species: Archae - 0; Bacteria - 19; Metazoa - 202; Fungi - 159; Plants - 80; Viruses - 10; Other Eukaryotes - 210 (source: NCBI BLink).	MGVIEDGTAKGCLPSEELFAVHYPAYPSSMSRA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
931	sp_156470_anam	-0.9100793	-0.8873636	-0.9441336	global transcription factor group B1 (GTB1); FUNCTIONS IN: transcription elongation regulator activity, hydrolase activity, acting on ester bonds, RNA binding; INVOLVED IN: chromatin assembly or disassembly, transcription initiation; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Resolvase, RNase H-like fold (InterPro:IPR006641), Nucleic acid-binding, OB-fold (InterPro:IPR012340), Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), Ribosomal protein S1, RNA-binding domain (InterPro:IPR003029), Transcription elongation factor Spt6 (InterPro:IPR017072); BEST Arabidopsis thaliana protein match is: Transcription elongation factor Spt6 (TAIR:AT1G63210.1); Has 59515 Blast hits to 32300 proteins in 2671 species: Archae - 146; Bacteria - 13496; Metazoa - 17051; Fungi - 6893; Plants - 5011; Viruses - 737; Other Eukaryotes - 16181 (source: NCBI BLink).	MAGRAVVSDEEEELEVGDDERTEVDEEEGEEEDD GGGGGGVDDDEDEDDEEGQDEFEDDGFIVEGD EDEEEPEEEDDEERHKKKKRKKKRESFELDDDDY ELLLDNNVKVARKADSKKFKRLKKARRDTDAGH SGLFDDDEFLGSGKGARTAQEQAKHTLFGDDDG GPVEDMPEEDEQPEADDVDLADEDDMADFIVSE EDEMDADGMPVRRMDKKKKKKHRQAAGVLPSSL QDAQDIFGDVEELLDMRRRANLEKDRYDENGLR KERNIGEEFEPTILAERYMTEKDDRIRDIDIPERM QISEESTGPPSTDAEAIVDEAKWIKKHLDSGQVP GLVNIASGEEMEYHIRNFLNFTHVQKLDLPYIAMY RKEDIFSLLKDPETGEDASHDIKPALKWHKVLWA VQDLDKKWLLLQKRKSALDGYYNKRFEEESRRIY DETRLRLNQQLFESIIGALRDSESEREVDDVDAK FNLHFPPGEAGVDEGQYKRPKRKSYYSLCNKAG LWEVASKFGHSSEQFGAAISLAVRVDEPEDPTET PEEVASNFTCAMFDSPQAVLKGARHMAAVEISC EPYVKKHFRNNYFDNVVVSTCPTPDGNVVIDAFH QFAGVKWLRNKPLNKFEDAQWLLIKKAEEEKLL QVTIKWPEEHLKKLIGDLELLLYTSLAVSKLAKPW NEQRRLILQDAIYNFILPALEKEARLLLTSRAKNW LLMDYGKCLWDKVSVAPYQRKESDLSSDEEAAP RVMACCWGYGKPPTTFVMLDSFGEVVDVLVAG SISLRSQNVTDQQRKKHDQQNLLKFMTEHQPQV VVLGAVSLSCTRLKDDIYEIVFKMVEENPRDVGH EMDGLSVFYGDESLPRLYENSRISSDQIPGHKGV VRRGVALGRYLQNPLAMVANLCGPGREILSWKIS SSESFLEDDEKYGMVEQIMVDVTNQVGIDLNLSA HHDWLFAPLQFISGLGPRKAATLQKSLVRAGAIV TRKDLLTAHGLGRKVFISAAGFLRIRRSGLAVSTS QFVDVLDDTRIHPESYALAQEMAKDIYREIVGDD NLDDDDAIDMAIENLRDRPSALKSFNVDAYAKDT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
932	sp_157110_zdeg	#N/A	-0.91876	-0.855241	protein arginine methyltransferase 10 (PRMT10); FUNCTIONS IN: histone-arginine N-methyltransferase activity, methyltransferase activity, protein-arginine omega-N asymmetric methyltransferase activity, [myelin basic protein]-arginine N-methyltransferase activity, protein-arginine omega-N monomethyltransferase activity; INVOLVED IN: vegetative to reproductive phase transition of meristem, peptidyl-arginine methylation, to asymmetrical-dimethyl arginine; LOCATED IN: cytoplasm; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Skb1 methyltransferase (InterPro:IPR007857); BEST Arabidopsis thaliana protein match is: protein arginine methyltransferase 6 (TAIR:AT3G20020.1); Has 3346 Blast hits to 3335 proteins in 890 species: Archae - 50; Bacteria - 1167; Metazoa - 1181; Fungi - 251; Plants - 354; Viruses - 1; Other Eukaryotes - 342 (source: NCBI BLink).	MGSPFANGGGERGLGGAANGSVDKGVDYANYF CTYSYLYHQKEMLSDRVRMDAYYNSIFQNTHHF QGKTVLDVGTGSGILAIWSAQAGARKVYAVEAT KMSEHARLLVKANGLEDVVEVIEGSMEDITLPEK VDVIISEWMGYFLLRESMFDSVICARDRWLKPTG VMYPSHARIWLAPIKSHLIDQKRDHFDGGMEEW HGFVDDTKNSYGVDMSILTESFTEEQKKYYLQTS LWNNLRPEQVIGTAAIIKEMDCLTATVDDILVIKTQ FFSSILEEDTRFSGFGGWFDVHFRGRSESPAKC EIELSTAPSVENGTHWGQQVFLLNSPIRSNKGDN MSISFVMNRSKENHRLLEAEFGCEFKLASGQPLQ PFRKKFFIE
933	sp_157330_kirt	#N/A	-0.855336	#N/A	RESURRECTION1 (RST1); FUNCTIONS IN: binding; INVOLVED IN: cuticle hydrocarbon biosynthetic process, embryo development ending in seed dormancy; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Armadillo-type fold (InterPro:IPR016024); Has 92 Blast hits to 88 proteins in 41 species: Archae - 0; Bacteria - 0; Metazoa - 50; Fungi - 0; Plants - 29; Viruses - 0; Other Eukaryotes - 13 (source: NCBI BLink).	MEAYAPLLVRTRVPQPSLQRYAVIQIFDKLRSTPS ILNSISDFEREAITQCLQSSSLAVVDQSVRELCRL VKDSKLDPSFALLELQSSLEGCPPRFVNVFVKGI GFLVRFGFCTRRFGSQNFVDAPENHPFIKVVTCR SEAQSELVQQVVLFVV-IGKQCGIAEVCDFLRPLL TFSVLRACSSDSSLSSFLSLVIYSLMSLCCSLLHE ANFLLEMLIGCLRCFNLRGAEDLAKVAIFAEFLVD AYSVVLRHIVKMGSMVTEVQLIGVKLLDALLTLCS DFEKFNIESKPVLELSRRLLSFQIELKLQYVPQFS SPMLSLCVLLAQLELEDEKLYVLDLLAYFLRWRL QDENVVAGAACNLSDLSLFFFPIISLMSSPSKTVN CSASQFLLMLEKLLVSHSTTSRDELSVQRGSSFV SKPETIIFRILQHLWLQDLSPCSFFLNINYASGNG NYGESGSKSWLSQLGKYALSFAERHKSCESISS SQEVMSTEISFLLGALAGSLVVHQSLGSYAIETLA ALNSMEPKLGVPLLLAVLYYCNLLRSNIIKYCDELL LKLLTMLSSLASNSAMLPLIYQTILPMLHKDSNPV LYATAIRLLCKAWEINDRVFGSLRSFLLPEYFVAS RFDRTIATSLAASLRDICRKSPERGVQLILSVAACI ECSVSTIRALGFQSLGYLCEADVVDFYTAWNVIA KYGQEYSTDPMIAYR
934	sp_158380_ikds	-0.87859	#N/A	#N/A	uncharacterized protein	MFFLHPAAFTLLLHLSARGQGREKAQLRDASMK GFCAVQSLLSPYHPPHYSARR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
935	sp_158590_ykrn	#N/A	#N/A	-0.866942	MAP kinase 6 (MPK6); FUNCTIONS IN: MAP kinase activity, kinase activity; INVOLVED IN: in 20 processes; LOCATED IN: trans-Golgi network, preprophase band, phragmoplast; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), JNK MAP kinase (InterPro:IPR008351), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), MAP kinase, conserved site (InterPro:IPR003527), Protein kinase, catalytic domain (InterPro:IPR000719); BEST Arabidopsis thaliana protein match is: mitogen-activated protein kinase 3 (TAIR:AT3G45640.1); Has 124968 Blast hits to 123484 proteins in 4565 species: Archae - 94; Bacteria - 13025; Metazoa - 47222; Fungi - 12634; Plants - 30512; Viruses - 574; Other Eukaryotes - 20907 (source: NCBI BLink).	MDALAPQPSDTDMMDAGAPPPENPLPPAADHS QTPPLQQQQQVSIPATLSHGGKFIQYNIFGNIFEV TAKYKPPIMPIGKGAYGIVCSSLNSDTNEHVAIKKI ANAFDNKVDAKRTLREIKLLRHMDHENVVAIRDII PPPRREAFNDVYIAYELMDTDLHQIIRSNQGLSEE HCQYFLYQILRGLKYIHSANVLHRDLKPSNLLLNA NCDLKICDFGLARVTSETDFMTEYVVTRWYRAP ELLLNSSDYTAAIDVWSVGCIFMELMDRKPLFPG RDHVHQLRLLMELIGTPSEHELGFLNENAKRYIR QLPAYRRQSLAEKFSHVHPSAIDLVEKMLTFDPR QRITVEDALAHPYLNSLHDISDEPICVSPFCFDFE QHALSEEQMRELIYREALAFNPEYHQE
936	sp_158850_ueym	-0.863148	-0.928812	-0.870484	nodulin MtN21 /EamA-like transporter family protein; LOCATED IN: membrane; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 7 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF6, transmembrane (InterPro:IPR000620); BEST Arabidopsis thaliana protein match is: nodulin MtN21 /EamA-like transporter family protein (TAIR:AT1G44800.1); Has 3616 Blast hits to 3605 proteins in 598 species: Archae - 32; Bacteria - 1860; Metazoa - 4; Fungi - 2; Plants - 1222; Viruses - 0; Other Eukaryotes - 496 (source: NCBI BLink).	MKDFAGTPGTITGYLIAAMGLQVIWSLSLALMDA YALAKKRVLHNSILVSLFVIGDWQLLHLQELRSCT SMISETIGWTLLYFITEASFNHGMNPYVYITYRHV VAGVVMFPFAYFLERKARPKLTVALIAEIFVLSLL GVGLALNMYFASLKYTSPTFVASMINTIASLTFVIA VALKMEVLDIWNPRGAAKVLGTLISLGGVLAMTL YKGSVIKSIWNPLLHIQGSTTTNENWLMGSILTTA SCLSWSIWYIMQAFTLKRYPALLSLTTWMSFVGA AQSAVFTACVQRKPAAWAIGFNIDLLSTVYAGVV CSGLIVFLQLWCTEEKGPVFVTMFNPLSTILVALL AYFVLGEKLYVGSILGAAIVIFGMYLLMWGKEEA QTNLAGIFRNYKDSRSLKPEAFTTTSEGREEITEP
937	sp_158941z6514	-0.874461	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
938	sp_159350_fmrp	#N/A	#N/A	-0.9883353	C-terminal domain phosphatase-like 1 (CPL1); FUNCTIONS IN: double-stranded RNA binding, phosphatase activity, nucleotide phosphatase activity; INVOLVED IN: abscisic acid mediated signaling pathway, negative regulation of transcription, response to abiotic stimulus; LOCATED IN: intracellular; CONTAINS InterPro DOMAIN/s: Double-stranded RNA-binding (InterPro:IPR001159), NLI interacting factor (InterPro:IPR004274); BEST Arabidopsis thaliana protein match is: carboxyl-terminal domain (ctd) phosphatase-like 2 (TAIR:AT5G01270.2); Has 330 Blast hits to 317 proteins in 105 species: Archae - 0; Bacteria - 31; Metazoa - 57; Fungi - 81; Plants - 112; Viruses - 1; Other Eukaryotes - 48 (source: NCBI BLink).	MMKTVVYEGDNLLGEVEIYFENNANKIEMMKKG MLMRISHYSEAERCPPLAVLHTITKSTGGVSFK MMEKSLYFQQHNDSQIFALHSSCLRGNKTAVVS LGEQEIHLVAMRSRRMDGVTTPCFWGFIVMPGL YESCLGMLNLRCLGIVFDLDETLIVANTLRSFEDRI EALQRKMTVEADPQRMAGMMAEVKRYQEDKAIL KQYAETDQVVDNGKVHKIQAEVIPALSDNHQTVV RPLIRLQDKNIVLTRINPQIRDTSVLVRLRPAWED LRSYLTARGRKRFEVYVCTMAERDYALEMWRLL DPDSNLIGGRELLDRIVCVKSGSRKSLFNVFQGGI CHPKMALVIDDRLKVWDEKDQPRVHVVPAFAPY YAPQAEANNAIPVLCVARNVACNVRGGFFKEFD EGLLQRMSDVYFEDDPKDFPSPPDVSNYLVSED DGSGSNANKEPICFDGMADAEVERRLKEAVLSS SSAPSLPCVTSLATVNLDHRLASSLPFSVAASSM TIPQPAPQASIAPFHANLFSQAGPLARTLASIGPK DLGLHSSPAREEGEVPESELDPDTRRRLLILQHG QDMREGLPNEPPFPGRPPVQAPVAGPGSGPGP VPVPGPVPVAGPGSASISVPGPGPVPMSGSVPA PAPVPVPPRVQSRGSWFQVEDHMNPSPLGRS ATKEFPMSPDAVHVEKQRPPPPFPRKVENPVWS DRSFPEKQRLPREASRRDERLRSNYSVPSHQSF RGDEISLSRSVSSNKGFEVEPEKGSSLSENPSVA LHDIAMRCGAKVEFKLGLVATSELKFFTEAYFVG EKIGEGTGTTRREAQYRAAEAALMNLADRYLTHI KSDASTPQSDTSRGPSPKDMGFASDANSQGDC TSKKEETTTPSSELTRLDDSILEGSKDSMGSVSV LKELCMIEGLGVEFKGQSPTSTNPVHGDEIHAEV EINGQVLGKGTGLTWDEAKMQAAELALASLKSM LGQITKRPSSPRLLQGMASKRLKPEYARVLEHMP SSRYPRNASPVP
939	sp_159420_duwo	#N/A	-0.8766127	#N/A	Arabidopsis protein of unknown function (DUF241); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: root; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF241, plant (InterPro:IPR004320); BEST Arabidopsis thaliana protein match is: Arabidopsis protein of unknown function (DUF241) (TAIR:AT2G17070.1); Has 750 Blast hits to 734 proteins in 87 species: Archae - 2; Bacteria - 37; Metazoa - 79; Fungi - 28; Plants - 534; Viruses - 2; Other Eukaryotes - 68 (source: NCBI BLink).	MASYHTRSFSFPSNSHPVADELDAQLSRLRSSQ AASTSSLTSKLSGLNDLYKCVEEFLQLPQNQKTV SQSQGENVVEQVLDGSLRLLDIVSTSRDVLALSK ERLQDIQSVLRRRCSGEIDITNEVSEYLKTRKASK KIIKKCLKDINAVDKTNESIAIESMLKDVQTLTVDIF KSLLSYISGSQKSSWSFSKLIRQGSDKEAATSISD FEAVDATLSQKNKANVDMSQLVKLESEIQEIDEV LEILFRHLIKTRATLLNVLSN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
940	sp_160190_qfxe	#N/A	-0.853413	#N/A	uncharacterized protein	MFGLGESSGSGRVRGGGRKGARAGARGSAHG GARGSPHITQSTQDTFVDDDMTQGEYDSEFVPE TQQDEEVVEPGVFWMTPGELWFDHSSLARHITN VIK
941	sp_160250_wrxf	#N/A	#N/A	-0.8502263	Peptidase S24/S26A/S26B/S26C family protein; FUNCTIONS IN: serine-type peptidase activity, peptidase activity; INVOLVED IN: proteolysis; LOCATED IN: integral to membrane, membrane; EXPRESSED IN: male gametophyte, pollen tube; EXPRESSED DURING: L mature pollen stage; CONTAINS InterPro DOMAIN/s: Peptidase S24/S26A/S26B/S26C, beta-ribbon domain (InterPro:IPR011056), Peptidase S24/S26A/S26B/S26C (InterPro:IPR015927), Peptidase S26A, signal peptidase I, conserved site (InterPro:IPR019758), Peptidase S26, conserved region (InterPro:IPR019533), Peptidase S26A, signal peptidase I (InterPro:IPR000223), Peptidase S26A, signal peptidase I, serine active site (InterPro:IPR019756); BEST Arabidopsis thaliana protein match is: thylakoid processing peptide (TAIR:AT2G30440.1); Has 9578 Blast hits to 9341 proteins in 2371 species: Archae - 0; Bacteria - 7160; Metazoa - 214; Fungi - 105; Plants - 244; Viruses - 0; Other Eukaryotes - 1855 (source: NCBI BLink).	MAIRYTVAFSATIAQNIASSATYRVTNYRCVHEFV SKPRPTVVNTPAISNPNLEPSKNYNHPYNCKNSQ PDARRSKPSFPPRQRTASFSTFAAELITGNSSSS SPVVSGLISLVNAAVFGGSISGSMGTMGVTPLRG ASIIPFLQVSKWLPCNEPALMRSTSSDVDRGGTS NILEESKIKKKVESRNSAAVDVSKVIAQQKVIDRK SWLSKMFNFSSDDAKAVFTAVTVNLLYKSSLAEP RSIPSSSMCPTLDVGDRILAEKVSYIFREPEVSDI VIFKAPPILQEIGFATGDVFIKRIVAKAGDCVEVKN GQLLVNGIVREEEFILEPLAYEMEPVIVPEGYVFV MGDNRNNSFDSHNWGPLPVKNIVGRSVFRYWP PTKLTDTIYDQHAEKFNVAVS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
942	sp_160320_mtxu	#N/A	-0.867471	-0.8868826	TBP-associated factor 5 (TAF5); FUNCTIONS IN: transcription regulator activity, nucleotide binding; INVOLVED IN: regulation of transcription; LOCATED IN: nucleus; EXPRESSED IN: guard cell, root, inflorescence, cultured cell, leaf; CONTAINS InterPro DOMAIN/s: WD40 repeat 2 (InterPro:IPR019782), WD40 repeat, conserved site (InterPro:IPR019775), WD40 repeat (InterPro:IPR001680), G-protein beta WD-40 repeat, region (InterPro:IPR020472), WD40 repeat-like-containing domain (InterPro:IPR011046), WD40-repeat-containing domain (InterPro:IPR017986), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), WD40 repeat, subgroup (InterPro:IPR019781), TFIID suounit, WD40-associated region (InterPro:IPR007582); BEST Arabidopsis thaliana protein match is: Transducin/WD40 repeat-like superfamily protein (TAIR:AT3G49660.1); Has 114463 Blast hits to 42274 proteins in 991 species: Archae - 68; Bacteria - 11258; Metazoa - 46869; Fungi - 25620; Plants - 15010; Viruses - 3; Other Eukaryotes - 15635 (source: NCBI BLink).	MEEEEIEKAVLAYLKKKGFRQTEVAFQEEQQLQ QQQKHPQQQQSKNSSSNSHTDPDVAKLILTFTS ESESLPAQYYDGYSKLRSWAHNSLDLYKHELLR VLYPVFIHCFMDLVARGHLQEARSFFITFREDHET MHERDLQKLEGVLSPSHLKEMEFAHTLRQSKVNI KICQYSHELLLQYLLKIQSMLGVVNEHINFQVCPG QPGIISDDADAVTLVDGGQDAANLINQKEIQWGM LEDSIEDRLEKGGGLLLDSDKVDGENKEGENEE NKKRSADGGKQAAPVKKSKKDKVGGAGGKAAR PEVITPLAPRVKPELTLPAMPTEVEHSILEDLRNR VQLSSVALPSVSFYTFINTHNGLNCASISHDGSLV AGGFSDSSLKVWDMSKLGQQMDNTALQGENDT AGNEHMHSANGGKIPYTLFQGHSGPVYSASFSP LGDFILSSSADSTVRLWSTKLNANLVCYKGHNYP VWDVQFSPYGHYFASCSHDRTARIWSMDRIQPL RIMAGHLSDVDCVRWHVNCNYIATGSSDKTVRL WDIQSGECVRIFIGHRSMVLSLAMSPDGRYMAS GDEDGAIMMWDLASGRCITPLVGHTSCVWSLAF SGEGSLLASGSADCTVKLWDVNASTKTPRSEEK TGSSNRLRSLRTLPTKSSPVYSLMFSRRNLLFAA GVLSKRS
943	sp_160380_zcsc	#N/A	#N/A	-0.8838414	unknown protein; Has 158 Blast hits to 158 proteins in 77 species: Archae - 0; Bacteria - 0; Metazoa - 104; Fungi - 0; Plants - 33; Viruses - 0; Other Eukaryotes - 21 (source: NCBI BLink).	MARRNHTTPLGCIACDELSELGAGTEFWLINNPS ILSALDTHSIALSNKSLILIFGWSDDPFDNIKVKIHP ELSPIDAEYITAIEWLVFDDIRVLAVGISSGYLLIYS LDGGLIHKQIYPTRILKLRVRGTKRDISQDISNEE VCIVMNGVIARFDGSDLQDMLRRWFRETQTNFW DQKRKKRDLDDFEDSFEERVPYQLWSVNKYGAC ADAAITGVMFPPLMDFEVLPY
944	sp_160490_zngu	#N/A	-0.8736385	#N/A	Protein of unknown function; DUF599; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF599 (InterPro:IPR006747); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF599 (TAIR:AT5G24600.2); Has 303 Blast hits to 303 proteins in 73 species: Archae - 0; Bacteria - 107; Metazoa - 0; Fungi - 0; Plants - 182; Viruses - 0; Other Eukaryotes - 14 (source: NCBI BLink).	YSQPSPVEFAAIQDKNFAIFNSQSYLLNPPRTDLE FQKIKISETSSSDPFYTNLRRNLHSHASFYTNLIAD ASSSLIRSFAGKVSCSAGKVSSARKTTFYYAHISF LLNVPESSDEGSEEFIEYVSQQLNRGCLFWSLGL RAFYFAFPVLLWIFGAIAMFACSCSMAILLYFLDT ASSSPSYLYNGNVVKTNRSPDIESRRVH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
945	sp_160690_zgoh	#N/A	-0.861183	#N/A	villin-like 1 (VLN1); CONTAINS InterPro DOMAIN/s: Gelsolin (InterPro:IPR007122), Villin headpiece (InterPro:IPR003128), Gelsolin domain (InterPro:IPR007123); BEST Arabidopsis thaliana protein match is: villin 4 (TAIR:AT4G30160.1); Has 3095 Blast hits to 1746 proteins in 186 species: Archae - 0; Bacteria - 6; Metazoa - 2441; Fungi - 102; Plants - 219; Viruses - 0; Other Eukaryotes - 327 (source: NCBI BLink).	GLEIWCVENLSLVPVPKSSHGKLFSGSVYLILNTV LLKDGMLEHDVHYWLGKDANEVESLMVSDKALE LDLALGSCTVQYQEKQGQESAKFLSYFKPCIIPC EGVYTSGPERSNRGTYHITLLACKGEHAISVKEV PFSRSSLNHDDVFILDTESKVYLFSGHNSSIQERA KALEVIKYIKEEKHGGKCQVATIEDGKFVGDSDA GEFWSIFGGYAPIPRDSSSFFKKQHDVPFWKLFLI TIQGKLCETSSGTFSREMLDREKCYILDCDAEVY VWMGRCTSISERKTSITATEDFLRSQGRSTRSLV MVLTEGSETPKFKSNFVGWPQTDNPKLYEEGKG KVAAIFKQQGYNIKELPEEYCQPLINCRGKLEVW GVNGDHLSIMPSEGDIKLFSGDCHIVKYTYQSSQ RDESILYAWLGTASIKEDFVDAISHIINMAKSTKG TPVLAQIFEGKEPDQFFWVTQRLIVLKGGTFTRYI NFIQEKGFADNTYSARSSALFRVQGSSLENMQAI QVNMVAGSLNSSYCYILQTEAAVFSWIGSLTLSR DHDLLDRMLELLNPAWQPKSFREGGEPDEFWS ALGGKVDYPREKEIRKHIEDPHLFVCNCSEDDFK VKEIFSFTQDDLTTEDVFILDCYTEIYVWLGSQSN VQPNLQTIQFGLKYLEGDILAQGLSLETPLYMVTE RHEPEFFTCFFDWDVSKSHMDGNSFERKLALLR GEKKLETPPRNSSRTYSSEYYQSGSRSKSVSST GMPRSASPSLSGSQLASSTAIAKKLLSEQPVNHS STVDFHKLESRNPADGADLSLVAENDDDDDANL LKYPYERLTLTSDEPVSDIDLTKREAYLSKAEFLE KFGMTTRAFYALPKWRQNKLKISLNLF
946	sp_161180_kmky	#N/A	-0.855802	#N/A	uncharacterized protein	MVVTVEARKFSLPYLSLSSPQVTSFHYDPNSLSL ALMHSDSTFSLFPSFSPSICNSLPPPQTLIPPPS SSACFVPLNANPNPKVQDGVSPDDKKRVVFVTA GPHNGGSKVVLRFHLLGIDGKFVSATQVDCSQN GLCFDRKLGVIVDVSHGMKVVLSCGSVNYLALYSA SGGKVWVFGVKLVGNGRGLRLVKCAVIDCRLRI SSIYLSPGFLILGEMRGVRIFPLRAIMVKGKVGSGR RLRVAKGKGEKSNVEGRDSKVPNGIIRPFAESNE LNKNTGNSNVSKLLPNGHACSVLDGVSSSSNAD GAVKSRSLKFRQDSGEWGMQFVLFNDCVHEGIK SGVMQSNKRKATSIQAISPEKFLLLDSNGDLHILH VSKSGSSRMRRLLTCTIQVQQLVAFPDASADSQIV WISDGLYSVHVMMIPGSDTKEKESGVSNDEREII QSSVTQAIFTSESIQHVVPLAANRILILGQDSLYAY GIS
947	sp_161260_iwdi	#N/A	-0.859783	-0.8453443	uncharacterized protein	MNDHEAH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
948	sp_161340_pmdg	-0.880334	#N/A	-0.900997	unknown protein; Has 158 Blast hits to 158 proteins in 77 species: Archae - 0; Bacteria - 0; Metazoa - 104; Fungi - 0; Plants - 33; Viruses - 0; Other Eukaryotes - 21 (source: NCBI BLink).	ASLTCLKDPPRKGEKLTLSPSGTLAAITDSLGRIL LLDTQALVVVRLWKGYRDANCLFMEMLVKKDSA STSYHEYKKSDYCLCLAIHAPRKGIIEVWQMRTG PRLLTIQCAKGNACPNKVQSHMHMLVERLNINPD QTQENNSSTLDPNVPIPGKWIKVTPIKKSAPRTR KPVLKLDQGGIIHLTPQNQVSQICVSTPKHQTPA ANQIQLHSNMYDQLLATENFKDMLREASDHEALL LENTMSAMSSSKASEFSSPLFEYGSNLPLGTGLS PKEEREFF
949	sp_162000_idqy	#N/A	#N/A	-0.867431	Protein kinase superfamily protein; FUNCTIONS IN: protein serine/threonine kinase activity, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: cellular_component unknown; EXPRESSED IN: sperm cell, male gametophyte, pollen tube; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Protein kinase, catalytic domain (InterPro:IPR000719), Serine/threonine-protein kinase domain (InterPro:IPR002290), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271); BEST Arabidopsis thaliana protein match is: Protein kinase superfamily protein (TAIR:AT5G45430.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MERYKLIKEVGDGTFGCVWRAINKQSGEVVAIKK MKKKYYSWEECINLREVKSLRKMSHPNIVKLKEV IREHDILHFVFEYMECNLYQLMKDRGRPFSENDV RNWCFQVFQALAYMHQRGYFHRDLKPENLLVSK DVIKIADFGLAREVCSGPPYTEYVSTRWYRAPEV LLQSPTYTSAVDMWAMGAIMAELFTLRPLFPGLS EADEIYKICSVIGTPTENSWGHGLLLSRAMNYQF PQLPGVHLATLIPSASDDAISLIRSLCSWDPIQRP TAAEVLQHPFFKNCFYVPPSLRTRATVNRTPPSL AARGYFDQKSARKSSGPGTLPAPKATNLKPHVA FGASVQRKLEMDGKDVNKNDKSIKSSAVKQPKY RPPARNSPSSIYPGRTSRGVSETAEKLSNMSLG SGRQTMRQSVPPPMKAGGWHGQSNMFMGRS QEIQPTRTFPRKVAG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
950	sp_162050_cedy	-0.336831	-0.940451	-0.886358	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors, oxidoreductase activity, iron ion binding; INVOLVED IN: oxidation reduction; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Isopenicillin N synthase (InterPro:IPR002283), Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (TAIR:AT3G19010.1); Has 8949 Blast hits to 8896 proteins in 1015 species: Archae - 0; Bacteria - 1163; Metazoa - 122; Fungi - 1047; Plants - 5005; Viruses - 0; Other Eukaryotes - 1612 (source: NCBI BLink).	MGDVNPDFIQPTEYRPNPTSTIEAKEIPVIDISSAS DEVIAEIKEACEKWGFFQVVNHGVPLKKFREIVEL AARKFFGLSMEEKKKVRRDEENPLGYYDSEHTK NVRDWKEIFDFTVQYPTLIPASHEIGDDEIKELRN QWPDYPPEFRVACEEYTQEMVKLSFKLLELIASS LGVPADRLSGYFKDHASFVRINYYPPCPYPDLAL GVGPHKDSGALTVLAQDDIGGLEVKRKMDGAWI LVKPIPDAYIINVGDVIQVWSNDKYESVEHRVKVN PDKERLSIPFFFNPSHYVMVKPLNELIDKQNPAKY REYNYGKFYTTRRLSDFKKLNVENIQVLHFRIQE
951	sp_162340_oprm	#N/A	-0.871097	#N/A	plant U-box 38 (PUB38); FUNCTIONS IN: ubiquitin-protein ligase activity; INVOLVED IN: protein ubiquitination; LOCATED IN: ubiquitin ligase complex; EXPRESSED IN: stem; CONTAINS InterPro DOMAIN/s: U box domain (InterPro:IPR003613), Armadillo-like helical (InterPro:IPR011989), Armadillo (InterPro:IPR000225), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein with ARM repeat domain (TAIR:AT5G62560.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGGNGKFLKLNFSFYRNSHSRSSSKAQSPPPPP PEYICPICNSLFTDPVVVGSGHTVDRLCAEVCEV HNFKPDFPDGTRPDFSTKIPNLALRTAIPTWCSK NGVALPQEKENVAVERIVKELLEKEERENGLIRP SERFLIDSVDKADPDLSHAVTELTHRANRFDTS TSSEESVIAAASPLTPLPFTTRPASCWSPFHTSS SDTLVSDEAMNPNPNFSDDEEFISKFKSCDPYEQ EQGAISLRKATRNDEEARVSVCTPRLLLALGSLLF SSYEKVQINAIAALVNLSLEKSNKLKIVRSGIVPTVI QLLRGRLSEAQEHAAGLIFSLSLEEKNRTAIGVLG ALPPLLNCLLRSDSERTRSDAALALYNLSLDHSN RVKLVKQFNAVPPLLAVLRQEKESIAGRVLLVLY QLAASTEGKAAMLDGNAVEHLVWMLRKGKGEF ESESTRENCVAALYALSHGSMRFKALAKEARAV EVLKEVEENGSERAKEKARRILVVLKGRIESEDG LGEGIDWEAILEPDGGLSRTRNRVAVGPGFGKT TEF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
952	sp_162460_epug	-0.940661	-0.97416	-0.950909	PIN domain-like family protein; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF652 (InterPro:IPR006984), Nucleotide binding protein, PINc (InterPro:IPR006596); BEST Arabidopsis thaliana protein match is: PIN domain-like family protein (TAIR:AT1G26530.1); Has 697 Blast hits to 697 proteins in 237 species: Archae - 32; Bacteria - 0; Metazoa - 241; Fungi - 212; Plants - 105; Viruses - 0; Other Eukaryotes - 107 (source: NCBI BLink).	MGRSKKAPKFAVRKKLVTSKTLKQHKQEVLNPK KKDLEAANLPRNVPNVSSALFFKYNTALGPPYRV IVDTNFINFSIQNKLDLEKAMMDCLYAKCTPCITD CVMAELEKLGQKYRVALRIAKDPRFDRLPCTHRG TYADDCIVDRVNQHKCYIVATCDRDLKRRIRKIPG VPIMYITQHKYSIERLPEATMGGAPRY
953	sp_162750_qchk	#N/A	-0.906544	#N/A	RING/U-box superfamily protein; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type (InterPro:IPR001841), Zinc finger, C3HC4 RING-type (InterPro:IPR018957), Zinc finger, PHD-type (InterPro:IPR001965); BEST Arabidopsis thaliana protein match is: RING/U-box superfamily protein (TAIR:AT2G47700.1); Has 1560 Blast hits to 1269 proteins in 195 species: Archae - 0; Bacteria - 91; Metazoa - 627; Fungi - 277; Plants - 288; Viruses - 4; Other Eukaryotes - 273 (source: NCBI BLink).	MGLDETDLIDDGDGGGKSLSSSNVSCSICLEPVT DNGGRSWAKLQCGHEFHLDCIGSAFNIKGAMQC PNCRKIEKGQWLYATGCRSFTDIMDDWPHDEDL YDLNYSEMFGVQWCPLPFSRLPLLEEGEFPSTA YHDFLGQQAVFAEHTAVSSSTHQCPYVAYYGPI HSSSSNSVGSVSDGSNYSNRWGGPSGPNEMP NSYAFTGMDPHYHGWEQHSLSFSASGTRLGAS DQPSAPSLTPRSSRVNSDLSRSGSFVHPFVINHS SGGRVGSSVTSSMIPPYPGSAARARDRVQALQA YFQPPSSSPPPMWAPIMSSTRRSSGHRGMSQG AAPMPSSSDQSSGGFFVYPPGSSTRSFPEPENP TSSRFHGWERDHLASFPLNQGDRDSSWLPHQV VGGGSSSFRQRHGSERTPQNRS
954	sp_162840_gkqt	#N/A	-0.906751	-0.925104	EMBRYO DEFECTIVE 2766 (EMB2766); FUNCTIONS IN: structural constituent of nuclear pore; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast, nuclear pore; CONTAINS InterPro DOMAIN/s: Nucleoporin, Nsp1-like, C-terminal (InterPro:IPR007758); Has 235342 Blast hits to 95277 proteins in 3388 species: Archae - 819; Bacteria - 58190; Metazoa - 65863; Fungi - 46438; Plants - 10106; Viruses - 2707; Other Eukaryotes - 51219 (source: NCBI BLink).	MAQFTGFSFNNTSSPSTSPFANPSSTAASSSTF GFGAPSSSPAFGAPTSSSAFGASSSSSASPSPA FGFGAASSSATPSFSLGSAASAPSFGFGSSSTTS VPGFGFGSSTTPSSSAFGSLPSSTAAPAFSFSLP ASASTTTTTSAVDPFPSSAPNLFGSSTSAPSTNT TLFGANNNTTTTPTFGSSSPATPPPLFASASSAA ATPSSLFASASSAAATPSSLFASASSSASPSSLF ASASSSLAASSSGFSFLKSSTQTATTTASPPSRF SFTPPASVAPASAAASTTSLFNFNNAASSTAGAS ATKPLFGANATPTTPPSVAASSLFSTPPTTTTTTS TPSATVSSPFAATTPAASSQPAATSTAPAAFSGV SLNSSTVTTPVASASPASTSFSGFGISSGSSSLT PTASFSSFTVPAVKPTTIPASSSQSQPQQTAATIA TTGGGAATTTTTTTTTTAPATGFLASTSVSAAISS STTTAVQTSSAVAVGSSSGTTSAVITPVTSAPTP KLPSEITGKTVEEIIKEWNTELQDRTRKFQKQAN AIADWDKRILQNRDVLLRLETEVAKVVETQSNLE RQLELIETHQQEVDKSLESIEEEAERIYKDERNSL LDDEAASTRDAMYDQAESVERELEHMTEQIKSII QTLNASQGGELDTTEGMSPLDVVVRILNNQLSSL MWIDEKAEEFSFRIQNLASEGSAADRELVAPRL WS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
955	sp_163080_qqgu	-0.991163	-0.971752	-0.958704	cleavage stimulating factor 64 (CSTF64); CONTAINS InterPro DOMAIN/s: RNA recognition motif. RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT4G00830.3); Has 21430 Blast hits to 18533 proteins in 865 species: Archae - 10; Bacteria - 1688; Metazoa - 9678; Fungi - 3064; Plants - 4656; Viruses - 0; Other Eukaryotes - 2334 (source: NCBI BLink).	MASTQQHRCVFVGNIPYDATEEQLIQICEEVGPV VSFRLVIDRDTGKPKGYGFCEYKDEETSLSARRN LQGYEINGRQLRVDFAENDKGADRNREQGRGG PGLATNVDAQKQVGSPATLVDASFHQPIGLPIAM TASSVMAGALGGSQTGKVVNQNGFQNQSLIGND PLTHHLANMSRSQLNQVIHEIKMLSMQNKDLARQ LLLVSPQLPKALFQAQIILGMVTPQMLQMPNLRQ QTPNPAVPTPRPSMQHTLQGHRPPVQSLTGLPS FTPLSSVSQNPLAHNQFSAPLPLHPQLRPQQPNL NQMVAPSQSTVPAHSGLSSHLSTRPQPLSGLHM PPVSSSSFRPVQAPPLQHPGHAGIKTTQNVNLQ PTFQSHSDPGYKVASSVRPASVEAIKKDTPVRAK TARDTTLASRTNTLKNLASVASVSDHKDSQDQM VRPSKVMKLNNERGTPFSSAGLATPSATAVASS SGAAGQSSSLEAQNIEKQAPQFPPDVDPALLQQ VLNLTPEQLSSLPPQQQLQVIQLQQMLR
956	sp_163340_tswk	#N/A	-0.961593	-0.958378	Homeodomain-like superfamily protein; CONTAINS InterPro DOMAIN/s: Homeodomain-like (InterPro:IPR009057), Myb, DNA-binding (InterPro:IPR014778), HTH transcriptional regulator, Myb-type, DNA-binding (InterPro:IPR017930), Myb-like DNA-binding domain, SHAQKYF class (InterPro:IPR006447), Homeodomain-related (InterPro:IPR012287); BEST Arabidopsis thaliana protein match is: Homeodomain-like superfamily protein (TAIR:AT3G46640.2); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEDQWEMGLPLPEDLTPISQSLISPGLALAFGITP DNTRAQFDLNLHDNCSTYSPSEENLESLSLRLFS GRSDFDCSNRMRRGDSFVNDEVNSGSKDDGDE NGLNRAVKRPRLVWTPQLHKRFVDVIEHLGIDKA VPKTIMEMMNVEGLTRENVASHLQKYRLYLKRM EGLSSEGQSNPRRDTPMFDCDGFGSMTMPHPH PPPPQQQQQCNSLPLMPMPMVVPPPGAAYGFG HLGLPLPSSSPSGYCGLESGAYNMFWNRQRG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
957	sp_163450_kfm	-0.906868	-0.915252	-0.956016	uridine kinase/uracil phosphoribosyltransferase 1 (UK/UPRT1); FUNCTIONS IN: uridine kinase activity, uracil phosphoribosyltransferase activity, kinase activity, ATP binding; INVOLVED IN: biosynthetic process, nucleoside metabolic process, metabolic process; LOCATED IN: cytoplasm; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Phosphoribulokinase/uridine kinase (InterPro:IPR006083), Phosphoribosyltransferase (InterPro:IPR000836), Uridine kinase (InterPro:IPR000764); BEST Arabidopsis thaliana protein match is: uridine kinase-like 2 (TAIR:AT3G27190.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MPEATTAIDYVMEKASGPHFSGLRLSPSTTSSPR ASASGSPPSASFFVDSDLPNQPFVIGVSGGTAS GKTTVCDMIIQQLHDHRVVLVNQDSFYRGLTDEE LECVHEYNFDHPDAFDTEQLLECVGKLQSGQPV QVPIYDFKTHQRRTDTFRQVNASDVIILEGILVFH DSRVRNLMSMKIFVDTDADVRLARRIRRDTVERG RDINSVLEQYAKFVKPAFDDFVLPSKKYADVIIPR GGDNHVAIDLIVQHISTKLGQHDLCKVYPNVYVIQ STFQIRGMHTLIRDQEITKHDFVFYSDRLIRLVVE HGLGHLPFTEKQVVTPTGSVYTGVDFCKKLCGV SIIRSGESMENALRACCKGIKIGKILIHRDGDNGK QLIYEKLPKDISERHVLLLDPVLATGNSACQAIELL IQKGVPEAHIIFLNLISAPEGVHCVCKRYPSLKIVT SEIDAALNEEYRVIPGMGEFGDRYFGTDD
958	sp_163460_kccu	-0.954412	-0.926955	-0.879334	MSFVTAQLVQPYRDRENEGSLSIYTEFSPMADPS FEAGRPGESAIELGRIIDRGLRESRAVDTESLCVL AGKLVWSVRVDLHIIDNGGNLLDAANVAALAALM TFRRPECTLGGVDGQEVIVHPPEVQEPRALIVHH LPIAVTFAFMTEDNILVVDPTYYEEAIMKGRFIATI NASEDVCAIQKAGEVGIAPTVVMQCLQIAAVKAA DITSKIESAVKLYNTERENRSVKRLSSVALLNISSP SSTMTDCKSNSICERTVVLQQMRKSPSRERHISE SNGMETEAPSCRGKEDASRLMKGPSCWDPFSK GVDSKHLKASLASCGKSMAIKENEKIVESKLLNLE VNSSEILKKRNSLASSDKATDVLPKSCEIKTLKDA VKPKSMRRKVISASNTR	

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
959	sp_164020_cmpp	#N/A	-0.917364	-0.937377	Magnesium transporter CorA-like family protein; FUNCTIONS IN: metal ion transmembrane transporter activity; INVOLVED IN: metal ion transport, transmembrane transport; LOCATED IN: membrane; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Mg2+ transporter protein, CorA-like (InterPro:IPR002523); BEST Arabidopsis thaliana protein match is: Magnesium transporter CorA-like family protein (TAIR:AT1G29830.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSGRSRDNSNSRRNAPTKPEVPKSYNNKDVIPG TELWVDGLICGFEYVRGRRKEQKQKTSSKFPYS RKLESEDSHKKTYAYGEADARDQNGIYNPDCSC HTELGSYPVSTLDGSMDNQAYKAEQYHGSYWV PIGWTRLFEIVQDVQVDVDWESKQFEFMETEDD VTVAELAAPYWERPVGPVWWCHVDASHQNIIS WLSNGQWLHPAVSIALRDESRLISERMKHLLYEV PVRVAGGLLFEILGQSAGDPLNEEDDIPVVLRSW QAQNFLVTAMHVKGSTSNVNVLGVMEVQELLIA GGYNAPKTVHEVIAHLACRLARWDDRFFRKSIFG AADEVELKFVNRRNHEDMHLFGAILNQEVRKLSR QVIRVKWSLHAREEITFELLHQLRGNAARSLLEMI RKSTREMLDEQEAVRGRLFTIQDVMQSNIRAWL QDRSLRVTHNLGVFGGCGLVLSIITGLFGINVDGI PGAENTPYAFGLFTGILLLIGAILIGVGLIYLGLKQP VSEEQVEVRKMELQELVKMFQHEAETHAQVRQ ENPRQSLPPTARVQFPLDADYVIIQ
960	sp_164030_dxyq	-0.886753	-0.927331	-0.954034	YUCCA 8 (YUC8); FUNCTIONS IN: NADP or NADPH binding, oxidoreductase activity, monooxygenase activity, FAD binding, flavin-containing monooxygenase activity; INVOLVED IN: brassinosteroid mediated signaling pathway, auxin biosynthetic process; EXPRESSED IN: 11 plant structures; EXPRESSED DURING: 4 anthesis; CONTAINS InterPro DOMAIN/s: Pyridine nucleotide-disulphide oxidoreductase, class-II (InterPro:IPR000103), FAD-dependent pyridine nucleotide-disulphide oxidoreductase (InterPro:IPR013027), Flavin-containing monooxygenase-like (InterPro:IPR020946); BEST Arabidopsis thaliana protein match is: Flavin-binding monooxygenase family protein (TAIR:AT5G43890.1); Has 11557 Blast hits to 11539 proteins in 1628 species: Archae - 52; Bacteria - 6582; Metazoa - 787; Fungi - 1560; Plants - 663; Viruses - 0; Other Eukaryotes - 1913 (source: NCBI BLink).	MDQMFNLAPVVENGDDHRRCVWVNGPVIVGAG PSGLAVGACLRDQGVPFVVLERADCIASLWQRR TYDRLKLHLPKQFCQLPKMPFPEYFPEYPTKRQF IDYLESYATHFDIDPRFNECVQYAKYDETFGLWR VKTVGQGPNKPEVEYVCRWLVVATGENAECVV PDIDGLANFNGKVIHSSEYKSGESYRGKKVLVVG CGNSGMEVSLDLSNHNAQPTVVCRSSVHILPREI MGRSTFELAMLLMKWIPLWLVDKILLFFAWFLLG DMQKYGLKRPNIGPLELKNKQGKTPVLDIGALAKI RSGDIKVVSGIKSFNANSVELIDGETVEVDSVILAT GYRSNVPSWLQEGEFFNKNGFPKTPFPSAWKG KSGLYAVGFTRKGLAGASFDAMKIAQDIGQVWK HETKQMKLTAPRHRRCISTF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
961	sp_164130_mmma	-0.8818878	#N/A	#N/A	cell division cycle 48C (CDC48C); FUNCTIONS IN: nucleoside-triphosphatase activity, ATPase activity, nucleotide binding, ATP binding; INVOLVED IN: embryo development ending in seed dormancy; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: ATPase, AAA-type, core (InterPro:IPR003959), ATPase, AAA+ type, core (InterPro:IPR003593), ATPase, AAA-type, conserved site (InterPro:IPR003960); BEST Arabidopsis thaliana protein match is: ATPase, AAA-type, CDC48 protein (TAIR:AT5G03340.1); Has 65227 Blast hits to 39412 proteins in 3299 species: Archae - 1912; Bacteria - 27659; Metazoa - 8516; Fungi - 6329; Plants - 5583; Viruses - 58; Other Eukaryotes - 15170 (source: NCBI BLink).	MRKRRSSHVARSSYWNHSMLLRRLGSCNRYES VEDVVKDLCSRYSEYKRLSAAILTKNVQEALSEL NWKQNKDSGRRDVNVRQSTSTSSSSSSSFESN SMSAPTDGQVSNPKFDFMKSTLRETYCHGGFES DKTEESMKKKKEITNVELLEKEEENDAMEDIVMH ERIDGGTGRKDLDLEHGRDKVQGQPMFKDFGG MHDVLEELKYNVLIPLRYPELLNRLSQLHAGPPT GILLHGPPGCGKTMLAQAIANEANVPLYQIAAPEI VSGVSGQSEENIRDLFLKARRTSPSIIFIDEIDAIAE KRENLQKGMEGRIVTQLLKCMDNIYGLALVADPK SSNCESRPSLSPAVPEISVPERRQSHVLVIGATN RPNSLDPALRRPGRFSWEIELGVPDENARFEILS VLTHDLGVQEGLDLKKMARDTQGFVGADLEEVV RKASLHAMQTMINLKKAGDNKERALEHHDDNW WKNLHFMHGEVETLYPTMADFEVAVKRVQPSLK REGFSTVPNVKWEDVGGLGTLREEFQSCIINPIK YPEEYGAFREDLPAGFLLYGPPGCGKTLIAKAVA NEAGASFIYIKGPELMDKYVGESELAVRKIFSRAR TCSPCILFFDEVDALTANRGESSGVTDRVLKQLLI ELDGAEERRGVYVIGATNRPELIDRALLRPGRFEE LLYIPLPSPQERGLILRALARKRPIDTNVDLIEIAER TACDNFSGADVACLMRRASRVSHQEMQLSKQK GSGAILWINTSHFEKALVGMSPSVNKKQREFYD SLSRSFKPT	
962	sp_164330_upjd	-0.344782	-0.9238816	-0.917909	ENTH/VHS family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF618 (InterPro:IPR006903), RNA polymerase II, large subunit, CTD (InterPro:IPR006569), ENTH/VHS (InterPro:IPR008942); BEST Arabidopsis thaliana protein match is: ENTH/VHS family protein (TAIR:AT5G65180.1); Has 1003 Blast hits to 995 proteins in 197 species: Archae - 0; Bacteria - 21; Metazoa - 515; Fungi - 219; Plants - 159; Viruses - 1; Other Eukaryotes - 88 (source: NCBI BLink).	MGSTFNPQILVEKLTKLNNSQQSIETLSHWCIFH MNKAKLVETWDRQYHSSPREQRLAFLYLANDIL QNSRRKGAEFVAEFWRVLPEALRDVIRNGDDFG KNAARRLINIWDERKVFGSRGHILKEEFVARQPE NIVNKSGKQLSIKFRQSIGIMFDKIISDYETIYGGQL DEDALLSRCTNTVSCLEKLIKDNSSGISSGNFSG SGFAEEFQGQHAALRECIEQLTTLEASRASLVSH LRDALQEQELKLDQVRRQLQAAQLQSDQVANVS QQFVSPDSNIVVMEERVVKDDHGSLAPPGFTSG DRQANPVVYTRQVPVSDKSAAAAEDDSRKSAAA AVAAKLTASTASAEMLSHVLTSLVSGVAFGNSTD NDPHHSDKRPKLENEASSYASAPPPLPPQNQQP PSPLTPFPHPESLQQNQNQNANNSTPTNQQSAG GEQPPLPSSSPPPLPPLPPMPLYPMPPQFIQTAG SMSNVAYSYGATLQPPPLSGYPSVTGMSPYAAP PVNNPYQPFPGSESGFYSPSSSVPMAPISRQQT L	
963	sp_164840_tepj	-0.9274455	-0.9863035	-0.9886622	uncharacterized protein	MAVVEVEEVSPASALREKAVL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
964	sp_166290_yrrd	-0.3136845	-0.907311	#N/A	Mitochondrial substrate carrier family protein; FUNCTIONS IN: binding; INVOLVED IN: transport, mitochondrial transport; transmembrane transport; LOCATED IN: mitochondrial inner membrane, membrane; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Mitochondrial carrier protein (InterPro:IPR002067), Mitochondrial substrate carrier (InterPro:IPR001993), Mitochondrial substrate/solute carrier (InterPro:IPR018108); BEST Arabidopsis thaliana protein match is: Mitochondrial substrate carrier family protein (TAIR:AT2G30160.1); Has 27351 Blast hits to 14015 proteins in 457 species: Archae - 0; Bacteria - 6; Metazoa - 11451; Fungi - 7735; Plants - 5292; Viruses - 0; Other Eukaryotes - 2867 (source: NCBI BLink).	MATTTDATPKFHNHELRPLLSLEEQQQQDLTMD PSPSHTHHDGLLFWQFMVAGSIAGCAEHMAMFP VDTIKTRMQALGSCPIKTAGVRTALRSILQSDGAA GLYRGIAAMGLGAGPAHAVYFSMYEFCKKSFSG GNPNNHVAHAMSGVCATVMSDAVLTPMDMVKQ RLQLSGGNGGAYRGVWDCVKLVLKKEGFGAFY ASYRTTVLMNAPFTAVHFATYEAAKKGLTEISPN TADDERLVVHATAGAAAGALAAVITTPLDVVKTQ LQCQGVCGCDRFTSGSIRDVIKTIAKKDGYRGLM RGWAPRMMFHAPAAAICWSTYEAAKTFFQELND QKNSSSLT
965	sp_166920_sydp	#N/A	#N/A	-0.866432	ubiquitin-specific protease 25 (UBP25); FUNCTIONS IN: ubiquitin-specific protease activity, ubiquitin thiolesterase activity; INVOLVED IN: ubiquitin-dependent protein catabolic process; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site (InterPro:IPR018200), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2 (InterPro:IPR001394); BEST Arabidopsis thaliana protein match is: ubiquitin-specific protease 23 (TAIR:AT5G57990.1); Has 8659 Blast hits to 7314 proteins in 254 species: Archae - 0; Bacteria - 9; Metazoa - 4447; Fungi - 1449; Plants - 1041; Viruses - 9; Other Eukaryotes - 1704 (source: NCBI BLink).	MSWQPSLVHQKRKNGGPLGLKNLGNSCYLNSV LQCLTYTPPLANFCLKNQHSSFCDSSSEPDRKR DCPFCIIEKRIARSLSIDLALDTPSKIVSCLRIFAEH FKCGRQEDAHEFLRYVIDACHNTCLRLKKLQQQ RPRKITNGSGGESFSGKTIVMDIFGGSLQSQVKC LSCGTESNKVDEIMDICLEISNCSSLKDAMKRFFQ AEILDGNNKYKCERCKKLVAAKKQMSILQAPNVL VVQLKRFEGIFGGKIDRVIAFEEILVLSSFMCKAS QDPHPEYSLFATIVHSGFSPESGHYYAYIKDAMG RWYCCNDSYVTLSTQQEVLSEKAYILFFTRTGQR PVSVDTAAVSNGVKPHEMNGSNMAKSPRSSIPV KAINTKSSNAPSFVKSIPLASKNNDASAVLQNKFG GITSSLGKRISGSDNGITEVQKRETVGSNGKLGH SVCKQMNGNNIHSLQDNNGNCSNGEHIVVANSQ VLVNGDAHLKNGTLGPLNASLHENNGLKSKATA GKMLDNGVIHNGQNNECIDVSGPRGKLETRETA RKQLDNGAINNGQNHKCTNVSGKRNSEDRDSCI LLGEDAQSLARVEEFKKLLQKEAALVLQTCGWS DKVYEYMRLKKRVCVNACSSTDKNELKKMLIADA RTTFISQIPGSLKGSLIEHIRSFSSQK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
966	sp_167360_rkor	-0.8953	#N/A	#N/A	peroxisomal adenine nucleotide carrier 1 (PNC1); CONTAINS InterPro DOMAIN/s: Mitochondrial substrate carrier (InterPro:IPR001993), Mitochondrial substrate/solute carrier (InterPro:IPR018108); BEST Arabidopsis thaliana protein match is: peroxisomal adenine nucleotide carrier 2 (TAIR:AT5G27520.1); Has 7303 Blast hits to 6606 proteins in 315 species: Archae - 0; Bacteria - 0; Metazoa - 2896; Fungi - 2326; Plants - 1372; Viruses - 0; Other Eukaryotes - 709 (source: NCBI BLink).	MALDFESLTEATSGAVGALVSTTVLYPLDTCKTK YQAEARSDHQQRYRNLSDVLWEAIRNGQVLSLY QGLGTKNLQSVIQQFIYFYGYSFLKRLYLKKSNFK SMGTKANLIVAAAAGALTVIITQPLDTASSKMQTS DFGKAKGLWKTLTEGKWSGAFDGLGISLLLTAN PSIQYTVFDQLKNRLLKTNNTNSDLSPQALSAST AFLLGALSKTIATCLTYPAIRCKVVIQAAEREEEKL PKSEKARKTISRVFYSIWKNEGILGFFKGLHAQIL KTVLSSALLLMIKEKITKSTWVMMLALRKYLLTTS TRLKSA
967	sp_167710_zhhd	-0.8416687	#N/A	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: plasma membrane; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G20310.1); Has 66 Blast hits to 66 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 2; Plants - 64; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MDKEKENKPKSLFLKFLPKAASAVSFQNPHFSPG RDHHHHNHHANHHNNHQYSSKLKTHAHKGFSG PINYSVLPDEAREKPKKDGSFEAHEPTSPKVSCM GQIKYKHTKNKLFRKIASKNNNNSLIFDNDLKKKTT KSSAISKFFNISKKKKSMGYPSEEHGKEGRRATLD NNNSKKKTNVPSLGMMNKFASGRESLASFDWT AQIAPIDHDRDYYSDDDQERRRDDDSDDEEERE VIIPYSAPMLIGGRSRGGGDDDDGDDNVMRTRK EVNLWKRRTMAPPPPLALNSKVRS
968	sp_168480_jwzx	-0.91534	-0.8366	-0.897363	Galactosyltransferase family protein; FUNCTIONS IN: transferase activity, transferring hexosyl groups, transferase activity, transferring glycosyl groups; INVOLVED IN: protein amino acid glycosylation; LOCATED IN: membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 31 (InterPro:IPR002659); BEST Arabidopsis thaliana protein match is: Galactosyltransferase family protein (TAIR:AT4G32120.1); Has 600 Blast hits to 599 proteins in 74 species: Archae - 0; Bacteria - 0; Metazoa - 168; Fungi - 0; Plants - 429; Viruses - 0; Other Eukaryotes - 3 (source: NCBI BLink).	MEGSLPLTTAAKGGERRGVGRSSASSYTSATTK PSLLLAFLSCVACLYVAGRLWQDAENRTLLVNLL QKNSDQRPKVLTVEDKLMVLGCKDLEKRIAEAE MDLTLARSQGYLRNQLRHESSTDKKLLAVIGVYT GFGSRLKRNTFRGSWMPRGDNLKKLEEKGIVIR FVVGRSPNRGDSLDRGIDEENRQTKDFLIIEGHE EAQEELPKIKAKFFFSTAVQNWDAEFYVKVDDNID LDLDGLIGLLESRRGQQSAYIGCMKSGEVVAEEG KQWYEPDWWKFGDQKSYFRHASSSLLILSKSLS EYIFVNSVSLKTYAHDDTSVGSWMMGLQATYIDD NRLCCSTNRHDKVCSVA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
969	sp_169340_gtah	#N/A	#N/A	0.6877833	cytochrome c oxidase 17 (COX17); FUNCTIONS IN: copper chaperone activity; INVOLVED IN: response to bacterium, response to copper ion; LOCATED IN: mitochondrial intermembrane space; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome c oxicase copper chaperone (InterPro:IPR007745); BEST Arabidopsis thaliana protein match is: Cytochrome C oxidase copper chaperone (COX17) (TAIR:AT1G53030.1); Has 294 Blast hits to 294 proteins in 132 species: Archae - 0; Bacteria - 0; Metazoa - 138; Fungi - 51; Plants - 55; Viruses - 0; Other Eukaryotes - 50 (source: NCBI BLink).	MRGVTSASVGVPAEQKNPETVTKATDQEMVTKA SEQDQKPKKKICCACPDTKKLRDECIVQNGEDAC GKWIEAHKQCLRAEGFNV
970	sp_169550_esgj	#N/A	#N/A	-0.9300311	uncharacterized protein	PRQSDATREAERRAMESEIKKRKINPPPRRVVSS FSATKEAFHKVNDSSMAASTPQSIQQARNHAIAQ AQQEGCTTNFRIFDSPFGNFLIPVIPTRAELSG
971	sp_169750_peix	#N/A	#N/A	-0.6722315	sorting nexin 2B (SNX2b); FUNCTIONS IN: phospholipid binding; INVOLVED IN: vesicle-mediated transport, intracellular signaling pathway; LOCATED IN: membrane; EXPRESSED IN: cultured cell; CONTAINS InterPro DOMAIN/s: Vps5 C-terminal (InterPro:IPR015404), Phox-like (InterPro:IPR001683); BEST Arabidopsis thaliana protein match is: sorting nexin 2A (TAIR:AT5G58440.1); Has 2399 Blast hits to 2386 proteins in 279 species: Archae - 13; Bacteria - 100; Metazoa - 1366; Fungi - 555; Plants - 146; Viruses - 0; Other Eukaryotes - 219 (source: NCBI BLink).	MMGSENEEGAHPLAPSDEMETLILDDPLSSSKSF ANYKSALISLSDSHNHPLSPSIVAKSVGSDPLLGP LHEEDQIRPKEVDFSSHFEDSDDYSDVFDSPITG GIGGGGNNGVGSPKVNSRSFSRSESLGSEYLNI TVLNPQKEHEVSSNSIVPGGNTYVTYLISTKTNM RGFGGSDFSVRRRFKDVVTLSDRLSESYRGFFIA PRPDKSVVESQVMHKQEFVEQRRGALEKYLRKL AAHPVIRTSDELRVFLQVQGKLPLPTSTDVASRM LDGAVNLPKQLFGDSRSAVAPHEAVQPAKGGRD LLRLFKELKQSVTNDWGNSRPAAVEEDKEFLEK KEKLFDFENRLTDVSKQAESLVKAQQDMGDTMG ELGLSFIKLTKFENEAAILGTQRTRATEMKNIATAT VRGSRMYRELTAKTVKNLDTLHEYLGVMLAAHN AYSDRANALLTVQTLLSEVSSLQSRAEKLEAASS RVFGSDKTRNHKLEELKETMKITEDSKNCAVREY ERIKENNRTELERLEREKHTDFLKMLKGFVNNQV NSAEKMANLWSKVAEDTRGYVKGSS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
972	sp_169770_doag	-0.86828	-0.913349	#N/A	auxin response factor 2 (ARF2); FUNCTIONS IN: protein binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: in 7 processes; LOCATED IN: nucleus, chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aux/IAA-ARF-dimerisation (InterPro:IPR011525), Transcriptional factor B3 (InterPro:IPR003340), AUX/IAA protein (InterPro:IPR003311), Auxin response factor (InterPro:IPR010525); BEST Arabidopsis thaliana protein match is: auxin response factor 1 (TAIR:AT1G59750.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MSSSEVASMKATDEAGIKSRHQHHHHGGAITGK VDPETALYKELWHACAGPLVTVPRPNELVYYFP QGHIEQVEASTNQAADKQMPNYSLPWKILCRVID VQLKAEADTDEVYAQVMLLPQPQREESSMEKEP PPPPQPHFHVVHSFCKTLTASDTSTHGGYDEVYA QVMWLPQPQREESSIEKEPPPPPQPYFHVHSFC KTLTASGMSTHGGFSVLRRHADKCLPPLDMSRQ PPTQELVAKDLHSNEWRFRHIFR
973	sp_169880_deqi	#N/A	-0.859407	-0.912916	root FNR 2 (RFNR2); FUNCTIONS IN: oxidoreductase activity, copper ion binding; INVOLVED IN: oxidation reduction; LOCATED IN: chloroplast, thylakoid membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Oxidoreductase FAD/NAD(P)-binding (InterPro:IPR001433), Ferredoxin reductase-type FAD-binding domain (InterPro:IPR017927), Riboflavin synthase-like beta-barrel (InterPro:IPR017938), Ferredoxin Reductase (InterPro:IPR015701), Flavoprotein pyridine nucleotide cytochrome reductase (InterPro:IPR001709), Ferredoxin--NADP reductase (InterPro:IPR012146); BEST Arabidopsis thaliana protein match is: root FNR 1 (TAIR:AT4G05390.1); Has 6645 Blast hits to 6645 proteins in 1680 species: Archae - 12; Bacteria - 3566; Metazoa - 779; Fungi - 733; Plants - 567; Viruses - 0; Other Eukaryotes - 988 (source: NCBI BLink).	MAHSAIAKVSVAVPVSNESAVKGSTFKTYSVSFG SKSWVPGYTLDLKTNNLRLRKQSVVCMSVQEAA ITKVAVSPLELEDAKEPPMHVYKAKGPYTGTIVSV ERVVGPKAPGETCHIVIDHAGNVPYWEGQSYGVI PPGENPKKPGNPHNVRLYSIASTRYGDYFDGKT ATLCVRRALYYDPETGKEDPAKMGVCSNFLCDS KPGDKVQIAGPSGKVMLLPEDPNATHIMIATGTG VAPYRGYLRRMFMEDVPAYKFKGLAWLFLGVAN TDSLLYDVEFSKYLKDHPDNFRYDTALSREHQNK KGGKMYVQDKIEEYSDEIFKLLDNGAHIYFCGLK GMMPGIQDTLKRVAEERGESWDEKLSQLKKNK QWHVEVY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
974	sp_170130_kxrs	#N/A	-0.877316	-0.3663482	Tetratricopeptide repeat (TPR)-like superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; EXPRESSED IN: fruit; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Tetratricopeptide repeat (TPR)-like superfamily protein (TAIR:AT4G36680.1); Has 20922 Blast hits to 6391 proteins in 198 species: Archae - 3; Bacteria - 20; Metazoa - 73; Fungi - 116; Plants - 20269; Viruses - 0; Other Eukaryotes - 441 (source: NCBI BLink).	LVANSRIEPKFEEPALIKLKKERDPDKLFKLFKDN SHNRLVIENKYAFEDTVSRLAGAKRYDYIEHLLEH QKSLPQGRREGFVVRVIMLYGRAGMVKQAVDTF YGMHLFGCRRTVKSLNAALKVLSQTHNLEAINSF LKEVPWEYGIQLDVISVNIVINAFCGMGILDKAYLV MVEMEKAGIKPDVVSYTTLISAFYKYGRWEVGN GLWNLMVYKRCYPNVATFNARIQFLVNKGRAW QANDLLRLMENIELQPDEVTYNLVIKGFFRAGLFD MAKVYSALHGRGYEPNLKIYQTMVHYLCKAGD FNCAYTMCKECLKKNWIMNVDTICTLLTGLKSTG KVEKANIIFSLVQKRHPPFPARQLDTFKAMLSRR
975	sp_170260_iqun	#N/A	-0.874119	#N/A	uncharacterized protein	LQIKQSEGEFGSGDVFYGGGGAGTGGGRAGM ASFRQQQQQQ
976	sp_170360_uufi	#N/A	#N/A	-0.3388284	CONSTITUTIVE EXPRESSER OF PR GENES 30 (CPR30); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: negative regulation of defense response; LOCATED IN: nucleus, cytoplasm; EXPRESSED IN: 8 plant structures; CONTAINS InterPro DOMAIN/s: F-box domain, cyclin-like (InterPro:IPR001810), F-box domain, Skp2-like (InterPro:IPR022364), F-box associated domain, type 1 (InterPro:IPR006527), F-box associated interaction domain (InterPro:IPR017451); BEST Arabidopsis thaliana protein match is: F-box associated ubiquitination effector family protein (TAIR:AT4G22390.1); Has 1743 Blast hits to 1730 proteins in 49 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 1741; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	RAHPAKIAGKTLIRFKVLSTFWHSTVSDPKFAKYY LGLTSRHHFLLFSPETKFLLLPYDEFNNLKDLEKI DIGGRRDVRFVGCCNGLVCLYNLNVLKSKRNYG SRYRRCGFLSICNPVTHQNREITNPGKSELALGY WFVYASSIDDYNIVALFSSRVFNYNEQYVHFCVF SMRTGIWKRYCHLDDYYDVYKNFSCKTTVMNDI MYWHPGCLGHEWEDKRICAFDLVNEEWKEYPW LDFLNEYFGADFFVRNGCLSLYCHKSHMVDDVW TLKHYDDRNSWEFLFSVNFGDWRPYTSIPTGKIL VRSLWLFKIFDVSNDRPEQGDARSLDYTLDIRAE SFVETLTSPFGTINFSKDDNEV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
977	sp_170680_hsig	#N/A	0.830332	#N/A	ABNORMAL SUSPENSOR 2 (SUS2); INVOLVED IN: nuclear mRNA splicing, via spliceosome; LOCATED IN: chloroplast, membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Mov34/MPN/PAD-1 (InterPro:IPR000555), Pre-mRNA-processing-splicing factor 8, U5-snRNA-binding (InterPro:IPR019581), Pre-mRNA-processing-splicing factor 8 (InterPro:IPR012591), PROCN (InterPro:IPR012592), Pre-mRNA-processing-splicing factor 8, U6-snRNA-binding (InterPro:IPR019580), PRP8 domain IV core (InterPro:IPR021983), PRO, C-terminal (InterPro:IPR012984), RNA recognition motif, spliceosomal PrP8 (InterPro:IPR019582); BEST Arabidopsis thaliana protein match is: Pre-mRNA-processing-splicing factor (TAIR:AT4G38780.1); Has 828 Blast hits to 729 proteins in 312 species: Archae - 0; Bacteria - 2; Metazoa - 333; Fungi - 220; Plants - 69; Viruses - 2; Other Eukaryotes - 202 (source: NCBI BLink).	WWNGNNPWAPPPAPGTSIPPPPAWGPSYTVLQ PPPQQPQQEESPAEAEARLEEKARKWMQLNSK RYGDKRKFGFVETQKEDMPPEHVRKIIRDHGDM SSKKFRHDKRVYLGALKFVPHAVYKLLENMPMP WEQVRDVKVLYHITGAITFVNEIPWVVEPIYLAQ WGTMWIMMRREKRDRRHFKRMRFPPFDDEEPP LDYADNVLDVDPLEPIQLELDEEEDSAVCTWFYD HKPLVKTKLINGPSYRRWHLSLPIMATLYRFAGQ LLSDMVDRNYFYLFDMESFFTAKALNMCIPGGPK FEPLYRDMEKGDEDWNEFNDINKLIIRSPLRTEY RIAFPHLYNNRPRKVRLCVYHTPMVMYIKSEDPD LPAFYYDPLIHPITSSHKDRRDKKAYEEDDDDFEL PEVVEPLLGDTQLYTDTTAAGISLLCAPRPFNMR SGRTRRAEDIPLVSEWYKEHCPPSYPVKVRVSY QKLLKCFVLNELHHIRPPKAQKKKHLFRSLAATKF FQSTELDWVEVGLQVCRQGYNMLNLLIHRKNLN YLHLDYNFNLKPVKTLTTKERKKSRFGNAFHLCR EILRLTKLVVDANVQFRLGNVDAFQLADGLQYIFS HVGQLTGMYRYKYRLMRQIRMCKDLKHLIYYRFN TGPVGKGPGCGFWAPMWRVWLFFLRGIVPLLE RWLGNLLARQFEGRHSKGVAKTVTKQRVESHF DLELRAAVMHDVLDAMPEGIKQNKARTILQHLSE AWRCWKANIPWKVPGLPVPIENMILRYVKAKAD WWTNVAHYNRERIRRGATVDKTVCRKNLGRLTR LWLKAEQERQHNYLKDGPYVTPEEAVAIYTTTVH WLESRKFSPIPFPPLSYKHDTKLLILALERLKESY SVAVRLNQQQREELGLIEQAYDNPHEALSRIKRH LLTQRAFKEVGIEFMDLYSYLIPVYEIEPLEKITDA YLDQYLWYEGDKRHLFPNWVKPADTEPPPLLVY KWCQGINNLQGIWDTSDGQCVVMLQTKFEKFFE KIDLTMLNRLLRLILDHNIADYVTAKNNFASFVVQY YGLVLDLLILGLTRGSEIAGQQPQMPNEFMTFWDT

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
978	sp_170710_zfhk	-0.86159	-0.849352	#N/A	Y-family DNA polymerase H (POLH); CONTAINS InterPro DOMAIN/s: DNA polymerase eta (InterPro:IPR017061), DNA-repair protein, UmuC-like (InterPro:IPR001126), DNA-repair protein, UmuC-like, N-terminal (InterPro:IPR017963); BEST Arabidopsis thaliana protein match is: DNA-directed DNA polymerases (TAIR:AT5G44750.1); Has 8088 Blast hits to 8015 proteins in 2292 species: Archae - 210; Bacteria - 5635; Metazoa - 500; Fungi - 401; Plants - 163; Viruses - 3; Other Eukaryotes - 1176 (source: NCBI BLink).	MPVAKPASSESRVIAHLDMDCFYVQVEQRKQPQ LRGLPTAVVQYNEWKGGGLIAVSYEARKYGVKR SMRGDEAKSVCPEIQLVQVPVARGKADLSIYRDA GSEVISILAKKGRCERASIDEDDTEVKERVKIWLQ RSDADHNDKLLSCGALIVAELRKHVLKETEFTCS AGIAHNKMLAKLASGLNKPAQQTIVPSASVKGFL GPWPIKKMKQLGGKLGDSLQIDLGINTVGDLLQF SEQKLQERYGFNTGTWLWNIARGVNGEEVKGRI LPKSHGSGKTFPGPRALRSIDTVSHWLKQLCEEL SERLLADLDQNKRIAHTLTLHVSAYKSSDTDSQK SFPSKSCPLRYGTTKIQEDVLSIFHAGLREFMGTF GAKSDKNQNPGWGVTGLSVSASKIISVPQGTCSI KKFFPSQCAVSCTSSEDTNNELLHERMQFPVRD WNLDSVASIGREDNAVMHDISGLHQNEDKRKSQ EDQGEATSFSLLDKATTSCSLLDQTTQTCSLLGQ ITDGRTLQLGREGSQSCGNIDQDEPQKFLDKEPA CLAVSPAPKIWDREVKKSKPKALNQGTASILRFL KRPEQSCSPLKPEPNGTLKDTSVPTCSGGSSLQ SGENQGKRERWSYDVHEIDPTIMEELPADIQEEL RSWMRPLKQHQSAKRSSITSYFLPSKDSS
979	sp_170780_demq	-0.916336	-0.923589	-0.981075	sodium:hydrogen antiporter 1 (NHD1); FUNCTIONS IN: potassium:hydrogen antiporter activity, sodium:hydrogen antiporter activity; INVOLVED IN: sodium ion transport; LOCATED IN: integral to membrane, chloroplast; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Na+/H+ antiporter NhaD-like (InterPro:IPR018390), Divalent ion symporter (InterPro:IPR004680); BEST Arabidopsis thaliana protein match is: Na+/H+ antiporter 2 (TAIR:AT1G49810.1); Has 2659 Blast hits to 2653 proteins in 794 species: Archae - 96; Bacteria - 1787; Metazoa - 39; Fungi - 48; Plants - 76; Viruses - 0; Other Eukaryotes - 613 (source: NCBI BLink).	MVLLFVIGYTGIIFEESLAFNKSGVGLLMAVSLWA VRWIGAPSNEVAMSELEHATAEVSQIVFFLLGAM TIVEIVDAHQGFKLVTENITTRKPRTLLWVVGFVT FFLSAILDNLTSTIVMVSLLRKLVPPSEYRKILGAV VVIAANAGGAWTPIGDVTTTMLWIHGQISTLPTM KALIIPSVVSLAVPLALMSLTSEVNGKGLESSSVM ASEQMAPRGQLVFAVGIGALVFVPVFKSLTGLPP YMGMLLGLGVLWLLTDAIHYGESERQKLKVPQA LSRIDTQGALFFLGILLSVASLEAAGILRELANYLD AHIPNIELIASAIGVFSAIIDNVPLVAATMGMYDVT SFPKDSEFWQLVAYCAGTGGSMLIIGSAAGVAF MGMEKVDFFWYFKKVSGFAFAGYAAGIATYLAL HNLDISLPTLPTTLAHIPFLSG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
980	sp_170840_tgxn	-0.952634	-0.963794	-0.961699	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: N-terminal protein myristoylation; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF632 (InterPro:IPR006867); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF630 and DUF632) (TAIR:AT5G25590.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGCSQSKIENEEAVSRCKDRKVFMKSAVTARNA FAAAHCSYAMSLKNAGAALSDFAQGEVVNPLSL SSAPVGPTSAGVGPTMGAIPSQAPPPPPVQDL PPPPPPIPGMPQPLQRAVSMPEFAVKAPGAKLV AEDTIEEEDEEEEDEQEEEEEEVGRLRRRRDN KGGSVGGRMETTPSISNLGTANTNIVGPPPSPPP PDERAIRTPPPPPLPEVHKQTWDYFFNTEDIPGG SLRDDDDIRGEYDRTVYHQRAMRSSASASSITTA LPPSIEMVEEMVMPEKPPMGGGGRVMKKGKQV VGQGGSGGEGRRKTVNWNLVQIFSDLDDCFLKA SESAHHVSKMLEATRLHYHSNFADNRGHIDHSA RVMRVITWNRSFRGLPNAEEFKDDFDSEENETH ATVLDKLLAWEKKLYDEVKAGEIMKFDYQRKVAL LNKQKKRGTNPESLEKTKAAVSHLHTRYIVDMQS MDSTVSEISRLRDEQLYPKLVALVDAMATMWEA MQVHHENQLKIASHLRFIDISQSTKETIDHHERTI QLWAIVQEWNTHFEKMVNHQKEYIKALNSWLKL NLIPIESSLKEKVSSPPRPQNPPIQNLLLAWHDHL EKLPDEHAKSAIHNFAAVLQTIVQQQEDEIRKRAR VEESKRDLEKKQRHYNKWLSEYMQKKIPNDDGE PGSVEDEFVHKDEIMVEKEFVVENAMKKLEEDS EDYERHCTQVREKSLASLKTRLPELFRTLSEFAL SASDMYSNLKGLSQFSQSHNKHKESVDA
981	sp_170870_apfp	-0.834911	-0.946425	-0.91166		

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
982	sp_171060_mnxd	#N/A	#N/A	0.8533333	Galactose oxidase/kelch repeat superfamily protein; CONTAINS InterPro DOMAIN/s: Galactose oxidase/kelch, beta-propeller (InterPro:IPR011043), Kelch repeat type 1 (InterPro:IPR006652), Kelch related (InterPro:IPR013089), Kelch-type beta propeller (InterPro:IPR015915); BEST Arabidopsis thaliana protein match is: Galactose oxidase/kelch repeat superfamily protein (TAIR:AT1G14330.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MLEGISFSSRDLPNACGQEHNWMPDGFSLMETS NYKFIKSEYGDDNKLGNSIKLLKINKAGDMEIDFLN LSLSLVVDPSDSIDNGREMSDTSSLIHQIGRDNSI DCVARCSRSDYGSIALVNRNFRSLIEGKELYKLR RQMGIVEHWIYFSCSLLEWEVFDPIRLRWKHLPI MPSNDCFRHSDKESLAVGTELLLFGKGITSHLIYK YSLLTNTWSTGMEMNTPRCLFGSASLGEIAIVAG GCDFWGNVFSLAELYNSDTGMWTTLPSMNKAR KKCSAVFMDGKFYVIGGIGVDNSNPLTCGEVFDL ERLTWTEIPDMFPLRIVEPGAPGASAMSEAPPLV TVVNNELYTADYARKEIRKYNKSSNSWATVGSLP ERVVSMHGWGLAFRGCGDKLIVVGGPRGLEGG YIEVNSWAPSESPQQWTVLGRKPGGFVYNCTIM GC
983	sp_172480_keig	#N/A	-0.87 48486	#N/A	Uncharacterised protein family (UPF0041); INVOLVED IN: biological_process unknown; LOCATED IN: mitochondrion; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised protein family UPF0041 (InterPro:IPR005336); BEST Arabidopsis thaliana protein match is: Uncharacterised protein family (UPF0041) (TAIR:AT4G14695.1); Has 886 Blast hits to 886 proteins in 209 species: Archae - 0; Bacteria - 0; Metazoa - 387; Fungi - 234; Plants - 170; Viruses - 0; Other Eukaryotes - 95 (source: NCBI BLink).	MASSKFKAFWNHPAGLKTIHFWAPTFKWGLSIA NIADFTKPPENISYPQQFVVAASGLIWTRYGMVIT PRNWNLCSVNFAMSVTGFCQLSRKLRHDVSTEK EPAAAAAVVKE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
984	sp_172680_nxoz	-0.856018	#N/A	#N/A	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity, ATP binding, ATP-dependent helicase activity, nucleic acid binding; LOCATED IN: nucleolus; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, ATP-dependent, DEAD-box, conserved site (InterPro:IPR000629), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEA(D/H)-box RNA helicase family protein (TAIR:AT5G65900.1); Has 44310 Blast hits to 43105 proteins in 3082 species: Archae - 753; Bacteria - 22440; Metazoa - 6214; Fungi - 4682; Plants - 2500; Viruses - 12; Other Eukaryotes - 7709 (source: NCBI BLink).	IQARSIPPLLEGKDVLGAARTGSGKTLAFLVPSVE LLHQARFTSRNGVGVVIICPTRELAIQTHAVAKDL LKNHSQTLGLVIGGAGRRGEAERLAKGVNLLVAT PGRLLDHLQNTKGFVFKNLKCLVIDEADRILEANF EEEMRQILKILPKERQTALFSATQTKKVEDLARLS LSKPIYVDVDEGRRKVTNEGLQQGYCVIPSAQRL ILLYSFLKRNLSKKVMVFFSSCNSVKFHSELLRYI QIDCFDIHGKQKQTKRTTTFFDFCKAEKGILLCTD VAARGLDIPSVDWIVQYDPPDDPKEYIHRVGRTA RGEGAKGNALLFLIPEELQFLGYLKSEKVPVKEY EFDTKKLANVQSHLEKLVTNNYYLNKSAKEAYRA YILSYNSHSMKDVFNVHKLDLQAVATSFCFSSPP KVDIKLDSNASKFRKKSRKIEGRFNESNPYGKRG GDDSRQFVRY
985	sp_172700_hxfa	-0.862861	-0.93614	-0.945454	Phosphoenolpyruvate carboxylase family protein; FUNCTIONS IN: isocitrate lyase activity, catalytic activity; INVOLVED IN: metabolic process; LOCATED IN: chloroplast, chloroplast stroma; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Pyruvate/Phosphoenolpyruvate kinase, catalytic core (InterPro:IPR015813), Isocitrate lyase/phosphorylmutase, conserved site (InterPro:IPR018523), Isocitrate lyase/phosphorylmutase (InterPro:IPR000918); BEST Arabidopsis thaliana protein match is: Phosphoenolpyruvate carboxylase family protein (TAIR:AT1G77060.1); Has 8771 Blast hits to 8771 proteins in 1489 species: Archae - 119; Bacteria - 4900; Metazoa - 32; Fungi - 522; Plants - 172; Viruses - 0; Other Eukaryotes - 3026 (source: NCBI BLink).	MGSLEASAPLGRTTVHRLIAEEGSVLMPGIQDAF SAAICAKTGFKACFVSGFGVSAALLGLPDFGLLT TTEVVETVRRITAAAPDLCVVVDGDTGGGGPLN VQRFIKELIAAGAKGVFLEDQVWPKKCGHMRGK SVVPAEEHALKIAAAREAIGDSDFFLVARSDARAP HGLQEAIRRANLYREAGADATFVEAPANLEELKE VVRDVKGLRIANMIEGGKTPLHTPTEFKEMGFHLI AHSLSTIYATTKALVDIMKVLKDKGTTRDDLEKIVT FSEFNDMISLESWYETESKFKNFTPKSLES

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
986	sp_172750_ooff	-0.8553673	-0.8485555	#N/A	seuss (SEU); FUNCTIONS IN: protein binding, transcription cofactor activity, DNA binding, protein heterodimerization activity; INVOLVED IN: multicellular organismal development, gynoecium development, embryo development, ovule development, regulation of flower development; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: SEUSS-like 3 (TAIR:AT4G25515.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MVPSGPPTPIGGTQSVNPSLLRSNSGMLGGQG GPMPSQGGFPSIVSPRNQFNNVNLLGNVANVSS LLNQSYGNGGANPGISGAGNGPRGGLDAGAES DPLSGMGNGMGFNPSSSYGAANIVNPSSSGQS QGQQFPNSSSNQMLSDQAQSQQHEPQNFQHQ QQQFSGNQQQQQQYRSALGGVGSVKLEPQASG DQQASAQQQHQMQSLRGLNQVKLEPHQLQSMR GMGTVKMEPQHSEQSMFMHQQQQQQQQHQLM CQMSMPRTNTQAAAAAQLNLLQQQRYIQLQQQ QHQQQQQHQLMKGIPQQRSQMHQQFSPQNLPL RSPVKPVYEPGMCARRLTHYMYQQQHRPEDNNI DFWRKFVAEYFAPNAKKKWCVSMYGSGRQTTG VFPQDVWHCEICKRKPGRGFEATVEVLPRLFKIK YESGTLEELLYVDMPREYQNSSGQIVLDYAKAIQ ESVFEQLRVVRDGQLRIVFSPDLKICSWEFCARR HEELIPRRLLIPQVSQLGTAAQKYQAATQNSSLN MSVSDLQSNCNMFVASARQLAKTLEVPLVNDLG YTKRYVRCLQISEVVNSMKDLIDYSRSTGAGPME SLAKFPRRTGTPPPSMQGQSQQPEEQQQQQQQ PQQQHHQQQPSQPTMGPNSNSNDQGSVQAAG MQIASSNGANSVNNNNNSLNVASTSTSQSTIAGL LHQNSMNSRQQNPMNNANSPYGGGGSSVQMP SPGSSSTMQQQTQPNPSPFQSPTPSSNNPPQT TSHGAMAATTTNHNSANSPVMSNIHQPAISGDG DTIDTQSSVQKLIQEMMMSSQLNGGGMMGGIGS LGTDAKNVNGILPMSGNSALNGLVGNGPMNSSN SSMGGGGSLGGMGGAMGQASMVNGIRAAMGN NSVMMNGRVGMASMARDQSVNHQQQDLGNQM LSGLGAVNGFNSLQYDWKPSP
987	sp_173030_ochg	-0.8843553	#N/A	#N/A	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein; FUNCTIONS IN: methyltransferase activity; INVOLVED IN: metabolic process; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised protein family, methyltransferase, Williams-Beuren syndrome (InterPro:IPR022238), Methyltransferase type 11 (InterPro:IPR013216); Has 984 Blast hits to 969 proteins in 433 species: Archae - 16; Bacteria - 445; Metazoa - 145; Fungi - 145; Plants - 54; Viruses - 0; Other Eukaryotes - 179 (source: NCBI BLink).	MAARPELQAPPEIFYDQSEARKYTSSSRIIDIQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
988	sp_173590_qkpi	0.886181	0.897314	0.93307	glycine-rich protein; CONTAINS InterPro DOMAIN/s: Plant specific eukaryotic initiation factor 4B (InterPro:IPR010433); BEST Arabidopsis thaliana protein match is: eukaryotic translation initiation factor 4B1 (TAIR:AT3G26400.1); Has 13357 Blast hits to 7743 proteins in 694 species: Archae - 46; Bacteria - 1331; Metazoa - 4294; Fungi - 1200; Plants - 649; Viruses - 189; Other Eukaryotes - 5648 (source: NCBI BLink).	MAATVSSAWAKPGAWALDSEEHEDELKQPIMA QPVKPEPPAADFPSLAAAAAAKPKKKNKAQTLSL AEFSTYSAPKPKAFVAGQKLTHDEILTLPTGPRE RTAEELERPRLGGGFRNYGGGASDDSRVRVSD RGSSRDRDMGPSRADEADDWSKTKKFSAFGGG APERSDRREKVGFFNSSRADEVDNWASNKSFV AAPVDGARRERRMGFESYSNGSGPDSGSWSKP REERGSGGVEADTWGRRESVGSGNGNGRPRL NLQPRTLPVENNDNGGQSVIPTEVKVASPVLIFT PTEKVKGANPFGEARPREEVLAEKGKDWKEIDE KLEAVKIKEAGNGEKSEGFGRRGYGLGRSESGD EKTVKSWRKPESADAVSSPSGSSADTTENSANT ADSSAAESTEKVSEEN
989	sp_173630_tgmz	0.917337	#N/A	0.96301	polypyrimidine tract-binding protein 2 (PTB2); FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: pollen germination, regulation of RNA splicing, regulation of translation; LOCATED IN: nucleus, cytoplasmic mRNA processing body, cytoplasm; EXPRESSED IN: 31 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: polypyrimidine tract-binding protein 1 (TAIR:AT3G01150.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MASVSSQPQFRYTQPPSKVLHLRNLPWECTEEE LIELGKPFGKVVNTKCNVGANKNQAFIEFADLNQ AIAMISYYASSSEPAQIRGKTVYLQYSNRQEIVNN KTTADTAGNVLLVTIEGDEARQYSIEVLHLVFSAF GFVHKITTFEKTAGFQALIQFSDTETASSAKNALD SRSIPRYLLPEHVGPCSLRITYSAHTDLSVKFQSH RSRDYTNQYLPVAPSAIDATGQLSVGVDGKKME PESNVLLASIENMQYAVTLDVLNTVFSAFGPVQKI AMFDKNGGLQALIQYPDVQTAVVAKEALEGHCIY DGGFCKLHLSYSRHVDLSIKVNNDRSRDYTVPIH GGMIQSSILGQQPPGGGPMQGGNAPPPYNGGQ YPGPPPPSGGWAPPQHMHMQMPMQNYQYNM QPGHGATPPGPPGTMPPGPMHMQNPPGMPPY SQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
990	sp_174500_wkrw	-0.874112	#N/A	#N/A	DNAJ heat shock N-terminal domain-containing protein; FUNCTIONS IN: unfolded protein binding, heat shock protein binding, nucleic acid binding; INVOLVED IN: protein folding; LOCATED IN: cellular_component unknown; EXPRESSED IN: male gametophyte; EXPRESSED DURING: M germinated pollen stage; CONTAINS InterPro DOMAIN/s: Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), RNA recognition motif, RNP-1 (InterPro:IPR000504), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Heat shock protein DnaJ (InterPro:IPR003095), Heat shock protein DnaJ, conserved site (InterPro:IPR018253); BEST Arabidopsis thaliana protein match is: ARG1-like 2 (TAIR:AT1G59980.1); Has 16747 Blast hits to 16541 proteins in 2695 species: Archae - 126; Bacteria - 6883; Metazoa - 3195; Fungi - 1541; Plants - 1234; Viruses - 13; Other Eukaryotes - 3755 (source: NCBI BLink).	MFFLSPPFTERSQIPYSMLGLVPGEEASKLTEQEI SSAYRKKARESHPDKNINDPDALAKFQKLQLSYE ILKDPKTRKEFDDLLRVRREKKVRDNKKFDAK
991	sp_174830_eocm	-0.884134	-0.935132	-0.966776	YELLOW STRIPE like 3 (YSL3); CONTAINS InterPro DOMAIN/s: Oligopeptide transporter OPT superfamily (InterPro:IPR004813); BEST Arabidopsis thaliana protein match is: YELLOW STRIPE like 2 (TAIR:AT5G24380.1).	MGSDGEVREMEREDYEEQNRVQDEIKRVPPWQ KQITIRGIIASILIGVIYSVVMKLNLTTGLVPNLNVS AALLAFVLLRSWTSLLHKLGISTTPFTRQENTVVQ TCAVACYSISVGGGFGSYLLGLNKNTYEQVGVD TEGNYPNSYKEPRIGWMTGFLFVSSFVGLLALVP LRKIMIIDYKLSYPSGTATAVLINGFHTPKGDKIAK KQVRGFVKYFSLSFVWGFFQWFYTGGDQQGFS QFPTFGLKAWKNSFYFDFSMTYIGAGMICSHLVN LSLLLGAVLSWGVMWPLIGDLKGKWFLSSLPES SMKSLNGYKVFISISLILGDGLYNFLKILFFTIRSIH MRLKGKNLNSCPNDQKRSLDDLQRNEFFMRERI PIWLACSGYTFFSIISIIIPKMFPELKWYYVVAAYIL APSLSFCNAYGAGLTDMNMAYNYGKVALFVLSA LAGKDSGVVAGLVGCGLIKSIVSISSDLMHDFKTG HLTLTSPRSMLLSQAIGTAVGCVVAPLTFFLFYKA FDVGNPDGEYKAPYALIYRNMAILGVQGFSALPQ HCLQLCYAFFGFAVVANLVRDFSPKKIGKWVPLP MAMAVPFLVGAYFAIDMCVGSLVVFVWHRLNSQ KASLMVPAVASGLICGDGLWILPSSILALAKLKPP MCMTFQSS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
992	sp_174890_kiax	-0.356557	-0.992383	-0.957487	NRPB7; FUNCTIONS IN: DNA-directed RNA polymerase activity, RNA binding; INVOLVED IN: transcription; LOCATED IN: DNA-directed RNA polymerase II, core complex; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold (InterPro:IPR012340), Ribosomal protein S1, RNA-binding domain (InterPro:IPR003029), RNA polymerase Rpb7, N-terminal (InterPro:IPR005576); BEST Arabidopsis thaliana protein match is: RNA polymerase Rpb7-like, N-terminal domain (TAIR:AT4G14660.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MFFHIILERNMQLHPRHFGPRLRDKLVSKLMKDV EGTCSGRHGFVVAIMGIEHVGKGLIRDGTGFVTF PVKYKCVVFRPFKGEILEAVVTMVNKMGFFAEAG PVQVFVSNHLIPDDMEFQSADTPNYTTSDNSVKI QKDSEVRLKIIGTRVDATEIFCIGTIKDDFLGVISDP GAAS
993	sp_175130_zkyx	-0.883711	#N/A	#N/A	uncharacterized protein	MEESSKQSGNSHHFEDFLDEEGENDEVVEAVQA HVKKKPSTDHVLGARRYDRIMAAFKPICLKALTS EDTMNIVVEWIRLLDIRLDNIIAQRSQNGDSLTQS PSYVQNKRFSNVKTPTSGSKSLGNCSKRKEPSP PAEQKSHTQENSSSNTFYVKDPISRKKAHSTCEV SLDLQPTTNVAIVDQQVTCIPTMDLFGDFKQVGS AYGGEADLIFENAMSVGGIELLHKDGSIGNFTQM VGTNQESEQECGGHY
994	sp_175420_ayyz	-0.986563	-0.967722	-0.976616	EMBRYO DEFECTIVE 975 (EMB975); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: LATERAL ORGAN JUNCTION (TAIR:AT2G39230.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MLFFTQGVYHNLIVDKKHVNLSRVVKVNCNSAVH LRKPPKYLRYPRRTKLPPDFGVDTFIKPRNEVKD HSEMDVQIDWSSDDQVTRVDFSDDETEAEDGH DVWEADEIEAISSLFQGRIHQKPGKLNRERPLPL PLPHQIRPLKLPTTKKHVRISSPAALSRLSLTKQIY KDPVFLLKLAREIRSLPAEENVSLVLNKWGRYLR KGSLSLTIKELGHMGLPDRALLTFCWAQKQTHLF PDDRILATTVEILARTHDLKLPPDLEKFVNLSSRG VLEAIARGFIKGGSLNLALKFLSAARRGKRLLDSSI YVKLLLELGKNPDKEEFVLPLLEELGERDSLNIVSL QDCTAIMKVCVRLGNFDFVESLFDWFQQNRNQP SVVMYTTVIYSRLSDNKYREALALVWEMERLNYV LDMPAYRVVIKLFVALNDLPRAVRYFSKIKEAGFS PTYDIYRDMISIYLLSERLAKCKEVCKEVELAGYK LDANLKSYLLQRER

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
995	sp_175560_xezr	-0.91234	-0.948148	-0.919582	nuclear transport factor 2B (NTF2B); FUNCTIONS IN: protein transporter activity, Ran GTPase binding; INVOLVED IN: nucleocytoplasmic transport, protein import into nucleus; LOCATED IN: nuclear envelope, intracellular, nucleus, plasma membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Nuclear transport factor 2 (InterPro:IPR002075), Nuclear transport factor 2, Eukaryote (InterPro:IPR018222); BEST Arabidopsis thaliana protein match is: nuclear transport factor 2A (TAIR:AT1G27310.1); Has 1018 Blast hits to 1018 proteins in 227 species: Archae - 0; Bacteria - 0; Metazoa - 491; Fungi - 232; Plants - 204; Viruses - 0; Other Eukaryotes - 91 (source: NCBI BLink).	MDPDAVSKAFVEHYYSTFDANRPGLANLYQDSS MLTFEGQKIQGSQAITAKLNSLPFQQCQHHISTV DCQPSGTGGMLVFVSGTLQLAGEQHALKFSQM FHLMPTPQQSFYVQNDIFRLNYA
996	sp_176670_gxpo	-0.867217	-0.908	-0.8934673	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT2G33410.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MDSDQGKLFVGGISWETTEEKLKDYFGQYGDVY QVNIMRDKATGRPRGFGFVLFSDPSVLDRVLQD KHTIDGRPVEAKRAMSREEQQSSGRTGNNNGG GRSGGGGGNFKTKKIFVGGLPPTLTDEGFRQYF ENYGQVTDVVIMYDQNTERPRGFGFISFDSEDAV DRVLQNNFHDLNGKQVEVKRALPKDANPGMGS RGNGGGYNSSGGQNAYGNRMDSSSYMQSQNA GGGFPAGYGSSGYGAAAYGYGPASGGMGYGG YAGYGGANPGYGGPASAAAYGNPNVSNAQRSS WNAQAASGYGAAGYGSSAGPWAASGAGSGGPT GQSPSGASGYGNQGYGYGSYGGSDGYGSVGN RAGGAPASNAAGGGGEMQGGYGDANGHGGYG AGGYGAQANGPQMGYGGGYGGARQAQ
997	sp_176860_cqhg	-0.881023	-0.87747	#N/A	RING/U-box superfamily protein; FUNCTIONS IN: zinc ion binding; EXPRESSED IN: leaf; EXPRESSED DURING: LP.04 four leaves visible; CONTAINS InterPro DOMAIN/s: Zinc finger, RING-type, conserved site (InterPro:IPR017907), Zinc finger, RING-type (InterPro:IPR001841), Zinc finger, C3HC4 RING-type (InterPro:IPR018957); Has 1875 Blast hits to 1873 proteins in 229 species: Archae - 0; Bacteria - 0; Metazoa - 654; Fungi - 156; Plants - 687; Viruses - 149; Other Eukaryotes - 229 (source: NCBI BLink).	MEKHSSSMRRPPSNCPNNKQSKTNSISSSIKGK TCPICLSLISGHRPAVITACFHAYCLRCLLRWSNY KRNCPLCNSHFDSYLSPSSPSSSSFHNHRLLLPLS LRRSPDPPRPQRRHAAVVQRLLRRRRDEIHAVN RQSRPVPRRRSFGLVDSAPSEVVTERILQWRASI YKHGISAVPFSGRNLQKLNMSVSRESILHRIEPWI QRELQAILEDPDPTIIVHVATSLYLSSIEKKTEALS TSSADHKDAFLAPLRPFLHELTDAFWHELRCFAE SCFTVDTYDEVVEYNKRK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
998	sp_176960_help	-0.8520112	-0.903586	-0.9377183	Calcium-binding EF hand family protein; FUNCTIONS IN: calcium ion binding; LOCATED IN: plasma membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: EF-HAND 2 (InterPro:IPR018249), EPS15 homology (EH) (InterPro:IPR000261), EF-hand-like domain (InterPro:IPR011992), Calcium-binding EF-hand (InterPro:IPR002048); BEST Arabidopsis thaliana protein match is: Calcium-binding EF hand family protein (TAIR:AT1G21630.1); Has 12017 Blast hits to 4944 proteins in 443 species: Archae - 20; Bacteria - 3590; Metazoa - 3650; Fungi - 1736; Plants - 412; Viruses - 16; Other Eukaryotes - 2593 (source: NCBI BLink).	MEVDTDRDGKITGEQARNLFLSWRLPREVLEQV WDLADQDNDSMLSLREFCIALYLMERYREGRPL PSSLPNNIMFDEALMRVPGLPNSSHGNSALGAT SGYRPVPGSLASSPVVPAAGLSPVMQPSLQPDM TGQFQQPKSRASVLDNSQANYFCEGWQPGIQE EASIWDEEWDKFEDEGKCRTCAFVVLEKGRK
999	sp_177160_edou	-0.8801026	-0.936375	-0.9233145	aconitase 3 (ACO3); FUNCTIONS IN: aconitate hydratase activity, copper ion binding, ATP binding; INVOLVED IN: in 6 processes; LOCATED IN: cytosol, mitochondrion, cell wall, plasma membrane, chloroplast; EXPRESSED IN: 27 plant structures; EXPRESSED DURING: 16 growth stages; CONTAINS InterPro DOMAIN/s: Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha (InterPro:IPR001030), Aconitase family, 4Fe-4S cluster binding site (InterPro:IPR018136), Aconitase A/isopropylmalate dehydratase small subunit, swivel (InterPro:IPR000573), Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha, suodomain 2 (InterPro:IPR015932), Aconitase/iron regulatory protein 2/2-methylisocitrate dehydratase (InterPro:IPR015934), Aconitase-like core (InterPro:IPR015937), Aconitase/3-isopropylmalate dehydratase, swivel (InterPro:IPR015928), Aconitase/iron regulatory protein 2 (InterPro:IPR006249), Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha, suodomain 1/3 (InterPro:IPR015931); BEST Arabidopsis thaliana protein match is: aconitase 1 (TAIR:AT4G35830.1); Has 20552 Blast hits to 20369 proteins in 2575 species: Archae - 558; Bacteria - 10777; Metazoa - 503; Fungi - 660; Plants - 232; Viruses - 0; Other Eukaryotes - 7822 (source: NCBI BLink).	MITGAATSSLSSSSISSLRAVRHRFASTLSLSPST NSSYSPTSRIFSAARSQPISGYRFSPGFRSICSSS SRWSHGVDWRSPVSLRAQIATATPTPLERFQRR IATMASEHPYKGIFTSLPKPGGGEFGKFYSLPAL NDPRVEKLPYSIRILLESAIRNCDGFQVTKNDVEKI IDWENSAPKLVEIPFKPARVLLQDFTGVPAVVDL AAMRDAMKNLNSDPNKINPLVPVDLVIDHSVQVD VARSENAVQANMELEFKRNGERFGFLKWGSTAF RNMLVVPPGSGIVHQVNLEYLGRVVFNNEGLLYP DSVVGTDSHTTMIDGLGVAGWGVGGIEAEATML GQPMSMVLPGVVGFKMTGKLRDGVTATDLVLTA TQMLRKHGVVGKFVEFHGQGVGQLSLADRATIA NMSPEYGATMGFFPVDHVTLEYLKLTGRSDETV TMIEQYLRANKMFVDYSEPQQERVYSSYLQLDLS EVEPCISGPKRPHDCVPLKEMKADWNACLNNKV GFKGFAISKDEQEKAVKFSFNGQPAELKHGSVVI AAITSCTNTSNPSVMLGAALVAKKATELGLEVKP WVKTSLAPGSGVVTKYLLKSGLQKYLDQQGFSIV GYGCTTCIGNSGDLHESVSAAISENDIVAAAVLS GNRNFEGRVHPLTRANYLASPPLVVAYALAGTV NIDFDKEPIGTGKDGKSVYFRDIWPSNDEIAQVV QSSVLPDMFKSTYEAITKGNPMWNSLEVPTGTL YQWDPNSTYIHEPPYFKNMSLNPPGPFSVKDAY CLLNFGDSITTDHISPAGSIHKDSPAAKFLMERGV DRKDFNSYGSRRGNDEVMARGTFANIRIVNKHLK GEVGPKTVHVPSGEKLYVFDAAMKYKEAGHDTI VLAGAEYGSGSSRDWAAKGPMLQGVKAVIAKSF ERIHRSNLVGMGVIPLCFKAGEDAETLGLTGHER FSIELPSKVSEIKPGQDITVTTDSGKSFVCTLRFD TEVELAYFDHGGILPYVIRQLMKQQ

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1000	sp_177190_opod	-0.926432	-0.919481	-0.973171	NHL domain-containing protein; LOCATED IN: endomembrane system; EXPRESSED IN: 7 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: NHL repeat (InterPro:IPR001258), Six-bladed beta-propeller, TolB-like (InterPro:IPR011042); BEST Arabidopsis thaliana protein match is: NHL domain-containing protein (TAIR:AT1G70280.2); Has 2211 Blast hits to 1172 proteins in 193 species: Archae - 21; Bacteria - 928; Metazoa - 143; Fungi - 0; Plants - 261; Viruses - 0; Other Eukaryotes - 858 (source: NCBI BLink).	MVEKMSKEINTAAGETPELPVSLHQQQVIDENEA LIMESFSSNNSCCSFFPCFGGQPSPNRSSGLGG STSSPSMWERLQQTAEIENKWWAGPLQGLKKM REWSEIIAGPKWKTFIRRFNRGGKSATGSGKFQ YDPLSYALNFDDGPGQNGHLDEDGYFPDFSSRF ASIPKSSRSSIDLGKNGIPSLS
1001	sp_177210_cfrw	-0.885886	-0.916781	-0.927344	trehalose phosphate synthase (TPS10); FUNCTIONS IN: transferase activity, transferring glycosyl groups, trehalose-phosphatase activity; INVOLVED IN: trehalose biosynthetic process, metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: HAD-superfamily hydrolase, subfamily IIB (InterPro:IPR006379), Glycosyl transferase, family 20 (InterPro:IPR001830), Trehalose-phosphatase (InterPro:IPR003337); BEST Arabidopsis thaliana protein match is: trehalose-phosphatase/synthase 9 (TAIR:AT1G23870.1); Has 4634 Blast hits to 4564 proteins in 1018 species: Archae - 52; Bacteria - 2697; Metazoa - 162; Fungi - 655; Plants - 601; Viruses - 0; Other Eukaryotes - 467 (source: NCBI BLink).	MSSKSYTSFLDLAAGGYLDLPQTPRNLPKVMTV PGVIPDADSTESKDDILDASTSRYRDRKIVVSNFL PINAQKDPKSGKWSFSLDDDSLYLHLKDGFSPES EVFYVGSLKVDVDVKEQEEVAQILLENFRCVPTF LPADLQKKFYHGFCKQYLWPLFHYMLPMSPDHG SRFDRGLWQAYVSANKKFADKVVEILNPEDDLV WVHDYHLMLLPTFLRRRFYRVKLGFFLHSPFPSS EIYRTLPVRDEILKAFLNSDLVGFHTFDYARHFLS CCSRLLGLDYESKRGYLGIDYFGRTVFVKILPVGL QLRRLEVAMNHPSASVKVKEIQKQFEGKKIILGV DDMDIFKGISLKLLAAEHLLNQHPELQGKLVLIQIV NPARSTGKDVQEAKSETYEITKRINMKFGFPGYE PVILIDRPVPLYEKTAYYAMAECCIVNAVRDGLNL TPYKYVICRQGSPKIDEAMGVSPGETRSSMLVVS EFVGCSPSLSGAIRVNPWNIEAVADAMKMAISIS ELEKQLRHEKHYRYVSSHDVAYWSRSFVQDLER ACKDHYNKRCWGIGFGLSFRILALSPSFRKLSVE YVVSAYKKATKRAIFLDYDGTVVSDMSIVKTPSSE LISVLQNLCSDTKNTVFIVSGRGRDSLTEWLSPC DKLGIAAEHGYFLRWKQNCDWEISTVTTDDDWK KIAEPVMQLYTESTDGSTIESKESAIVWHHQDAD PDFGSCQAKELLDHLENVLSNEPVVVTRGHNIVE VKPQGVSKGLVTERVLSELTNDGKAPDFLLCIGD DLSDEDMFESILSKASCSTSTGSVPEIFACTVGQ KPSKAKYYLNDTKHVMRLLQGLTFASNGKPPECI PPCQVSFESFA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1002	sp_178000_xtsc	#N/A	-0.919958	-0.904872	GPI transamidase subunit PIG-U; INVOLVED IN: GPI anchor biosynthetic process; LOCATED IN: plasma membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: GPI transamidase subunit PIG-U (InterPro:IPR009600); BEST Arabidopsis thaliana protein match is: GPI transamidase subunit PIG-U (TAIR:AT1G12730.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MGEEDSTKSSTKTKEVTRFWEWAIAAIIFRLVLIY FFPDILSQLASRPEVSTPVTSLRRLAEGYWLKKQ ASISPYAGSMYHGSPLLLSVLGPLTVRNIDGQPS HILCSLLFVIADLIAALLLRATGQKLQSAYCQSLKA LGVEKILGGSEFYASGDISALFYLWNPLTILTCMG SSTSPIENFLVILSIYGACKRLIPLAAVGWVLATHI SLYPAVLLIPVILLLGYGPDAPPKKLFLQKISRETA KDSSDSSLQHVKPITGSALPSFSWRPVILFLLWA SAWSIYILVLCGISVRQNGGLQEMFKRTYGFILTV EDFSPNIGVLWYFFAEVFDFFRNFFLIVIHMNILFM ILPLAIRMKHRPCFLAFVYVAICSMLKTYPSVADS ALYLGLFGLFINQLADMHFSFVLFCGYIGVNLLSP VMHNLWIWRGTGNANFYYATAIAYAALQIILVVDS ASAMLNHDRRLRKLTTA
1003	sp_178120_ypai	-0.923755	-0.85193	-0.916327	FUC95A; FUNCTIONS IN: 1,2-alpha-L-fucosidase activity; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Six-hairpin glycosidase-like (InterPro:IPR008928), Uncharacterised conserved protein UCP007663 (InterPro:IPR016518); Has 1349 Blast hits to 1316 proteins in 300 species: Archae - 2; Bacteria - 1123; Metazoa - 0; Fungi - 116; Plants - 52; Viruses - 0; Other Eukaryotes - 56 (source: NCBI BLink).	MDPLQLPSLVLPLCIFLSFSLVILGRFGAMEEHND WVMGRKPWEKDIWKPSTSLVMGVEESRPLKIIF KGPAKNWTDALPIGNGRFGAMVHGGLSSEIINLN DDTLWTGVPGDYTDPNAPQAIANVRKLVDSGNY SEATAFADKELSGDSSNAYQLLGDLILEFDDHDS AYIEQTYHRELDLDTATVTVKYSVGDVEFTRQLF ASSPDQVIVAKISSSKQGSLSFTASLDSKLDHHSY TNEKNQIIMEGSCSGYPKGIQFSLVLNLQIDGESG KIHNLQGRDLRVEHADSAVILLVASSSFDGPFTKP SDSKKDPTSEALSRINSINNILYSTLYTRHLDDYQ KLFHRVSLQLSKNSNIDEKISTAQRVQSFQTDED PSLVELLFQYGRYLLISSSRPGTQPANLQGIWSK DVYPPWSCAPHLNINLQMNYWLSLSCNLQELQK PLFDYLSFLAINGKKTANVNYKASGWVAHQVTDI WAKTSPDSGQAVWALWPMGGAWLSTHLWEHY TYTMDKNFLENKAYPILEGAASFLLDWWIEDKDG YLESNPSTSPEHMFTSLNGKPASVSYSSTMDVAI TKEVFSIIISAAEELGRTRDDLVRNVLKAQAKLPP YRIAKDGSLMEWAQDFDDPDPHHRHVSHLFGIF PGHSITPEEEPDLCKAANYSLYKRGKEGPGWST VWKAALWAHLHSSEHAYQMVKHLFDLVDPERE GSYEGGVYSNLFTAHPPFQIDANLGFSAAIAEML VQSTVKDLYLLPALPRDKWSNGCVKGLRARGGV TVNICWTDGDLHEVGLWSNGSNSVKRLNYRGTT VMADILPGTVYTFDSQLKCVNTQSLL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1004	sp_178590_uaxz	-0.886773	-0.851279	#N/A	aldehyde dehydrogenase 3F1 (ALDH3F1); FUNCTIONS IN: 3-chloroallyl aldehyde dehydrogenase activity, aldehyde dehydrogenase (NAD) activity; INVOLVED IN: cellular aldehyde metabolic process, oxidation reduction, metabolic process; LOCATED IN: endoplasmic reticulum, membrane; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Aldehyde/histidinol dehydrogenase (InterPro:IPR016161), Aldehyde dehydrogenase (InterPro:IPR015590), Aldehyde dehydrogenase, N-terminal (InterPro:IPR016162), Aldehyde dehydrogenase NAD(P)-dependent (InterPro:IPR012394); BEST Arabidopsis thaliana protein match is: aldehyde dehydrogenase 3H1 (TAIR:AT1G44170.2); Has 51213 Blast hits to 51182 proteins in 2839 species: Archae - 470; Bacteria - 31264; Metazoa - 2305; Fungi - 2096; Plants - 995; Viruses - 0; Other Eukaryotes - 14083 (source: NCBI BLink).	VGPAKKAVVHALSCLKEWMTPKKVGMRLLLYPS KAEIVPDPYGLYLIMTSWNYPLDLALEPLIGAIAAG NVIVVKPSELAPATSSFLANTVPLYLDNTAIKVIEG GPEASQQLLQQKWDKIFYTGSPKIGRIVMTEAAK HLTPVTLELGGKCPLIFDYSSISSNLKVALRRLSS AKWGACCGQACISIDYVLVEEKFTSSLIDMLKTLV RRLYGENIETLGSISRIVNKRHFDRVCNYLKDPAV AASIVLGGSYDEKRLLIEPTILLNPPLDSAVMNEEI FGPILPIITVKNIQESIDFINTRPKPLALYAFTHNDA FKKRIVSETSSGSLVFNDVMIQYICEELPFGGVG QSGMGRYHGKYSFDAFSHEKSVLYRGYFPDLSA RYPPWNGFKMEFCRLGYKIDYINLLLLLLGLKGH PKDD
1005	sp_179380_zgug	#N/A	-0.83948	#N/A	ATSTE14B; CONTAINS InterPro DOMAIN/s: Isoprenylcysteine carboxyl methyltransferase (InterPro:IPR007269); BEST Arabidopsis thaliana protein match is: homolog of yeast STE14 A (TAIR:AT5G23320.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MSELFSDTSCRQLSQFIVAIIFFHVSEYILAVCSQG RSNVTLKSLLISKNYYVAMICSLLEYWLEITFFPGL KEYWWVSNSGLFMVVFGETIRKVAMVTAGHAFT HLIKTHHEEQHRLITHGIYGFVRHPGYCGFFIWSV GTQIMLLNPICTLAFAVVWKFFAGRIPYEEFFLR QFFDDEYEEYAQRVPSGVPFVK
1006	sp_180010_hgfn	#N/A	-0.886336	#N/A	uncharacterized protein	MADLLPSDFPRRCRISFHQTFHRPLSPMLPNFIC NLRRLSFTNPYSALTYWDTLNR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1007	sp_180150_fjqs	#N/A	-0.885786	#N/A	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: DNA helicase activity, nucleoside-triphosphatase activity, nucleotide binding, ATP binding; LOCATED IN: nucleolus, chloroplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: TIP49, C-terminal (InterPro:IPR010339), ATPase, AAA+ type, core (InterPro:IPR003593); BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT3G49830.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MAELKLSESKDLTRIERIGAHSHIRGLGLDSTLEP RLTSEGLVGQLPARKAAGVILQMIKDGKIAGRGV LIAGQPGTGKTAIAMGLAKSLGHETPFAMLSGSEI FSLEMSKTEALMQAFRRAIGVRIKEEAEVIEGEVV EVQMDRPAVAGAASKTGKLTLKTTDMETVYDLG TKMIEALSKEKVQSGDVIAIDKASGKITKLGRSFS RSREYDAVGPHTKFVQCPDGELQKRKEVVHCVT LHEIDVINSRTQGFLALFTGDTGEIRAEVREQIDT KVAEWREEGKAEIVPGVLFIDEVHMLDIECFSFLN RALENEMAPILVVATNRGHTRIRGTNYKSPHGIPI DLLDRLLIIATQPYSEEEIRKILEIRCEEEDVEMSE EAKALLTKIAVETSLRYSINLINTSALACLKRKAKS VEMEDVSRVYQLFLDVKRSTQYLMEYQNQYMFN DVPGGEVNEDENTDMTS
1008	sp_180170_rpok	#N/A	-0.902141	-0.896135	Nucleotide-sugar transporter family protein; FUNCTIONS IN: nucleotide-sugar transmembrane transporter activity, CMP-sialic acid transmembrane transporter activity; INVOLVED IN: carbohydrate transport, nucleotide-sugar transport; LOCATED IN: endomembrane system, integral to membrane, Golgi membrane; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Nucleotide-sugar transporter (InterPro:IPR007271), UDP/CMP-sugar transporter (InterPro:IPR021189), UDP-galactose transporter (InterPro:IPR004689); BEST Arabidopsis thaliana protein match is: Nucleotide-sugar transporter family protein (TAIR:AT4G35335.1); Has 1068 Blast hits to 1053 proteins in 185 species: Archae - 0; Bacteria - 3; Metazoa - 567; Fungi - 123; Plants - 182; Viruses - 0; Other Eukaryotes - 193 (source: NCBI BLink).	MQWYFVAALLTVLTSSQGILTTLSQSNGGYKYDY ATVPFLAEVFKLVVSSILLWREVQTSPSVRMTTD WKTVRLFPIPSVIYLIHNNVQFATLTYVDTSTYQIM GNLKIVTTGILFRLFLKKKLSNLQWLAIILLAVGTTT SQVKGCGESSCDSLLSAPIQGYMFGILSACLSAL AGVYTEFLMKKNNDSLYWQNVQLYTFGTIFNLAR LVLDDFTSGFEKGPWWQRLFDGYSMATWLVVL NLGSSGLLVSWIMKYADNIVKVYSTSMAMLLTMI LSVYLFDFKPTLQLFLGIVICMMSLHMYFSPANAL VESPAMAKPEHESLKEVAIERLNS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1009	sp_180270_gdyp	#N/A	-0.651323	#N/A	cold shock domain protein 1 (CSDP1); CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold (InterPro:IPR012340), Cold-shock conserved site (InterPro:IPR019844), Zinc finger, CCHC retroviral-type (InterPro:IPR013084), Cold shock protein (InterPro:IPR011129), Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), Zinc finger, CCHC-type (InterPro:IPR001878), Cold-shock protein, DNA-binding (InterPro:IPR002059); BEST Arabidopsis thaliana protein match is: cold shock domain protein 3 (TAIR:AT2G17870.1); Has 93964 Blast hits to 39618 proteins in 2725 species: Archae - 55; Bacteria - 19157; Metazoa - 4501; Fungi - 1919; Plants - 2885; Viruses - 60267; Other Eukaryotes - 5180 (source: NCBI BLink).	MAQESRCRGVVQRFNDTRGFGFIKPEDDGEDLF VHHSDIKSDGYRTLTAGATVEFTFSVGSDNKPKA VDVTGPDGAPILPPPAARRSSGNGGGGRQRNN YSDGGKCYNCGESGHIARDCEGGSDNLGGGSG GGGNCYNCGKSGHFARACQSGSSAGGGSNVGR GGRVSADGDGACFNCGDIGHMARDCGQDKVNS GGRYGGRAGGGGGACYNCGDIGHLARDCAQER ASSGGRFGGRAGGSSGACYTCGEPGHLARDCS QDRPSSGRRSGGVSGGRNCYQCGEPGHIARDC SNAVA
1010	sp_181010_dffa	-0.857613	#N/A	#N/A	atypical CYS HIS rich thioredoxin 4 (ACHT4); FUNCTIONS IN: oxidoreductase activity, acting on sulfur group of donors, disulfide as acceptor; INVOLVED IN: cell redox homeostasis; LOCATED IN: chloroplast membrane, chloroplast, chloroplast stroma; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Thioredoxin fold (InterPro:IPR012335), Thioredoxin, core (InterPro:IPR015467), Thioredoxin-like (InterPro:IPR017936), Thioredoxin domain (InterPro:IPR013766), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: atypical CYS HIS rich thioredoxin 3 (TAIR:AT2G33270.1); Has 2537 Blast hits to 2529 proteins in 580 species: Archae - 19; Bacteria - 606; Metazoa - 627; Fungi - 356; Plants - 522; Viruses - 3; Other Eukaryotes - 404 (source: NCBI BLink).	MADVLRKSIVFSPKTSHRNNHSKRKFFPIKCSFP ANNGGRLVGISSRSSFGGVGTAVDQNRNIWSQK SRNFQSIGAKMGLSIGKAHKWWEKGLQQNMREI TSAEDLVDSLLNAGDNLVVVDFFSPGCGGCRAL HPKLCQLAEMNPDVQFLQVNYEEHKSMCYSLNV HVLPFFRFYRGAEGRVCSFSCTNATIKKFKDALA KHGTDRSSLAPTKGLDEKELVALAANKELSFTYT SKPVEAEPVPEHQEVLVNADDKAAAKESLSSNR KTLGSSSTLRSAQENEDKTLATARR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1011	sp_182620_smwd	-0.877663	-0.955325	-0.920558	sequence-specific DNA binding transcription factors; CONTAINS InterPro DOMAIN/s: SANT, DNA-binding (InterPro:IPR001005); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G76870.1); Has 317 Blast hits to 297 proteins in 56 species: Archae - 2; Bacteria - 28; Metazoa - 67; Fungi - 2; Plants - 135; Viruses - 9; Other Eukaryotes - 74 (source: NCBI BLink).	MDNNLSLANMMPGGGSAYVGSDMQGSIRVNHH QNPRSLQLGQGLMTQPSSMHEVFPLKMGQGHK RDQPNSMADYNKGERGKTSPTDEDEHSSNGEG MDGQSDPSKGKKVPQWHRVKWNSSMVKLLITA VSYLSEDAVFEGCSQGRRKYLNLHKKGKWKSIS KVMAERGFFVSPQQCEDKFNDLNKRYKKLNEIL GKGTSCQVVEKPALLDMMDHLSDKTKDEVRKIL SSKHLFYEELCSYHNNNRLHLPHDPALQQSLQLA LRSRDDTDTDFRKHLHGEAIEDDDGEDSEDHED YGENCSKGAYPLSGDPSKRMKQGQGLENSDFR PQSTQDFSKSFQLQLQNPFADMHPEFPENQKAA LLQKQRLKAQSFQLEEQKLQINSEALELEKQRFK WQRFCKRKDRELEKLKIENDRMKLDNERLALELK RMEMSFTSNH
1012	sp_183351z8694	#N/A	-0.855085	#N/A	uncharacterized protein	#N/A
1013	sp_183530_yzwm	-0.851888	-0.854419	-0.888942	KOW domain-containing protein; FUNCTIONS IN: structural constituent of ribosome; INVOLVED IN: translation; LOCATED IN: ribosome, intracellular, large ribosomal subunit; CONTAINS InterPro DOMAIN/s: Translation protein SH3-like (InterPro:IPR008991), Ribosomal protein L24/L26, conserved site (InterPro:IPR005825), KOW (InterPro:IPR005824), Ribosomal protein L24 (InterPro:IPR003256); BEST Arabidopsis thaliana protein match is: Translation protein SH3-like family protein (TAIR:AT5G54600.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGWKAAQKLIRNWKILRGDNVMIIRGKDKGETG TVKRVVRSQNRVIVEGKNLVRFFLSL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1014	sp_184260_ofeo	#N/A	#N/A	-0.905991	ubiquitin-specific protease 2 (UBP2); FUNCTIONS IN: ubiquitin-specific protease activity; INVOLVED IN: ubiquitin-dependent protein catabolic process; LOCATED IN: proteasome complex, chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Zinc finger, UBP-type (InterPro:IPR001607), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2, conserved site (InterPro:IPR018200), Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2 (InterPro:IPR001394); BEST Arabidopsis thaliana protein match is: ubiquitin-specific protease 1 (TAIR:AT2G32780.1); Has 10400 Blast hits to 6714 proteins in 274 species: Archae - 0; Bacteria - 50; Metazoa - 5279; Fungi - 1927; Plants - 1459; Viruses - 5; Other Eukaryotes - 1680 (source: NCBI BLink).	MGKKVKKKAARTAHKEKEKEREREKEKEKEKRV STSSPNAVSLEDNQSSVTADDGPVLQKERDACV HLEKGIDLEKFSLKIESAESTGCEDCRDGVNDRK ACKGKGKHGKKKGASDSRANKATWICLQCGHF SCGAVGLPTSPLSHALRHSRQNRHPLVIQLENPL LRWCFSCNTLLPVDKLVDIGEHKDVFLDIVKLLKK QSSKVRSVDVEDVWFGSGSVVSEVKEESKAINL VNGKDFYSVRGLVNLGNTCFFNSVMQNLLAMDV LRECLLNLEGSVGPLTSALKKFFVESSPESGLRN VINPKSLFGCVCAKASQFRGYQQQDSHELLRCLL DGLSSEDLYAKKSAGSNGKSKPVSSTFVDAIFGG QTSSSVCCVECGHSSVVYEPFLDLSLSVPTKNPL FKKARPVHRAKKTKLPPKRGGRIRPKLNKNEDM GPAPSNCTSSTSIESSSGLKGGLKGDMGTAYSG CNTSTSCESSSSLFAEAPSCAPSADFIGPINLVSR NAMMTQDNNSVEPDDASTEMLQSDAQDADGMS WLTYVEPISQSDGSQGADCNSWMDYLGPISLSD DPNSAGKVGFSVTQDFGSQDGNHDIISIDGTSKP IVEANLSQDSTMEGDNLGDVTVPDTSKSGGEVS EQPQYVNLKAESTADDWGEERAVQVQENEILLL TYKENGPVDGEEMKVDAEISSSAIVDGEDSLGFD GLGDLFNEPEAAEAGMNNTPVSDGLGTGFLAGN STESDPDEVDDTNSPVSVESCLAHFIKPELLSGE HAWHCENCSKLVKEGRKRMRKNQQKTALIQRN KSSNVCSTSEILHMSHSFPDSLAGLGNGSSDTDE NLKKHDTRRTNELCQDLPQEEEIRCVPEAAHLSA SGPAVYSIPSDGKSSDTHDNGPVEKPSGLSEEH ESGEEEEEEVDLKCVKVMRDATKRILISKIPPILTI HLKRFSQDARGRLSKLNGHVGFKEYIDLGPYMD CRSTERDNCIYHLIGVVEHSGSMRGGHYVAYVR GGDRRKGRSEEETIRGRSVWYHTSDAYVRETTL EEVLGCEAYILFYEKAKLKL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1015	sp_185550_myud	-0.822826	-0.883647	-0.948433	DDB1-CUL4 associated factor 1 (DCAF1); FUNCTIONS IN: nucleotide binding; INVOLVED IN: in 6 processes; LOCATED IN: nucleus, CUL4 RING ubiquitin ligase complex; EXPRESSED IN: 29 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: WD40 repeat-like-containing domain (InterPro:IPR011046), LisH dimerisation motif, subgroup (InterPro:IPR013720), WD40-repeat-containing domain (InterPro:IPR017986), WD40 repeat (InterPro:IPR001680), WD40/YVTN repeat-like-containing domain (InterPro:IPR015943), LisH dimerisation motif (InterPro:IPR006594); Has 5754 Blast hits to 3482 proteins in 410 species: Archae - 46; Bacteria - 1277; Metazoa - 1225; Fungi - 769; Plants - 278; Viruses - 87; Other Eukaryotes - 2072 (source: NCBI BLink).	MMEATTSSVEEPPSSSQSQPQPSPSPPPPEQTED SAAGEEVEDDDEEDEEDDVSVSKAQKFMEIITSS PDNPNLKALHALASLLETQESRYMEEADHSSSN QTRASHSIGRLGNIVRENDEFFELISSKYLSESRY STSIQAASARVLLSCSVTWMYPHVFEDAVTENIK NWVLDDSARDSCGDGGLWKQQFGEKKCSDFD MRKTYSTGLLALGLTSGSQVVEDVLTSGLSAKLM RCLRNRVLGETSISQKDTTHVLDSKSASLGGGGI GRDESRARLRPSLEAASLDDSRIQEDGGSADLV GERDHDRSFNWQVHGEEQWINAEPPDGLAEEV DIHDADIDGEGCWRGRDVLDGRLKIGERNGPGR SLHDEDYNESTRDDSARRRVNRWMRSRGKTRV HESTIEPEQPIMSPGSGSRGGQLRSSRERSTTR NLDGKRGTDNKKSSGRTTAEGKAVDREDIDECIY ECRIGTKDISDLIKTAVRSAEAEARAVDAPIEAIKV AGEAAAELVKSAAMDEFAATNSEDAAFLAASRA ASTVIDAGHAANTDEVSRNSDNASVDLPKDGAA EPDQNDDIEFFIPDSDTLAQQMEKYCIQCLVLLGE YVEVLGPVLHEKGVDVSLALLQRYSKEKKESEIV SSLPEVLKLICALAAHRKFAAVFVDRGGIQKLLAV PRVLHTFFGLSSCLFTIGSLQGIMERVCALPSDVI RRLVELGLQLLECPQDQARKNAALFFAAAFVFRA ILDVFDAQDGLQKSLGLLLRDAASVRSGVNSGTLG ASAPGALRNDRSNADVLTSSEKQIAYHTCVALRQ YFRAQLLLLVDTVRPNKSNRSIARHIPSIRAGYKP LDISNEAMDAVFLLIQKDRKLGPALVRARWPAVE RFLACNGHVTMLELCQAPPVERYLHDLLQYALGI LHVTIVPYGRKLIVTATLSNDRVGIAVILDAASGP GYIDPDIIQQALNVLINLVCPPPSISNKSPLLAQNQ QSISVQTSNVSTETRDRNTERSVSDRSHPVPGQ NEARERNGEPSTAERGLSTNNIMQSPSPNIASGL VGDRRISLGAGAGCAGLAAQLEQGYRQAREAVR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MGSSSIVTAERSSPEVPPWLKTLPVAPETRPTIG
						EFQDPIAYIFKIEKEASKFGICKIVPPVPPQSKKSVI
						SNLNQSLLNSSKNPSNPQPTFTTRQQQVGFCPR
						KQRPIQKSVWQSGECYTLQQFEAKAKAFEKNYL
						KNRAKKAVSPLEVETLYWKANGDRPFSVEYAND
						MPGSGFMPMKERKRTGEIVAGAAANVGETAWN
						MRGVARENGSLLRPMKEEIPGVTSPMVYVAMMF
						SWFAWHVEDHDLHSLNYLHLGAGKTWYGVPKD
						AASAFEEVIRVYGFGEEINPLVTFAQLGEKTTVMS
						PEVLINAGVPCCRLVQNAGEFVVTFPRAYHSGFS
						HGFNCGEASNIATPEWLRFAKDAAIRRASINYPP
						MVSHFQLLYDLALAICSRASIGNSAEPRSSRLKDK
					relative of early flowering 6 (REF6); FUNCTIONS IN: sequence-specific DNA binding transcription factor activity, zinc ion binding, nucleic acid binding; INVOLVED IN: in 6 processes; LOCATED IN: nucleus; EXPRESSED IN: 8 plant structures; CONTAINS InterPro DOMAIN/s: Transcription factor jumonji/aspartyl beta-hydroxylase (InterPro:IPR003347), Zinc finger, C2H2-like (InterPro:IPR015880), Transcription factor jumonji (InterPro:IPR013129), Zinc finger, C2H2-type (InterPro:IPR007087), Transcription factor jumonji, JmjN (InterPro:IPR003349), Zinc finger, C2H2-type/integrase, DNA-binding (InterPro:IPR013087); BEST Arabidopsis thaliana protein match is: Zinc finger (C2H2 type) family protein / transcription factor jumonji (jmj) family protein (TAIR:AT5G04240.1); Has 105567 Blast hits to 35250 proteins in 1010 species: Archae - 0; Bacteria - 23; Metazoa - 99094; Fungi - 2247; Plants - 673; Viruses - 7; Other Eukaryotes - 3523 (source: NCBI BLink).	KKGEGEMLVKQMFVQDVIQNNELLYTLGQGSEV
						VLLPHNSSEIFVWSNLRVGSKYKVKPGLPFSLYS
						SEEAIKASDDIMLARDDRKQKAFSSVKTKSGGCM
						STTPHEIQHSETEKGGGAAGDGFSDRGLFSCVK
						CGIWTFACVAIVQPTESAAQYLMSADCNSFNDWI
						AGSGVSSHGMDPIDGEANISDPNSFSGSMEKHP
						PDGAYDIPVHSTDYHAQSISNTSKLKSNTAEIVEV
						GSHTETKRESSALGLLAMTYGNSSDSDEDDVQP
						NSPVISEDNLSGDGSWGARFHQDDSASPVFEQ
						GYDSGAERGPSQISSRSECEDEDSSQRSDFYEH
						CGHRRVNGDDNEYDSHNCSAKFTEEDILTSEQN
						YSPIADEHDTAKISCAIDPVGKPNLSFAHRCDEDS
						SRMHVFCLEHAVEVEKQLRPIGGVHILLLCHPDY
						PNVEVEAKKAAEELEMDYGWKDIAFSMATKEDE
						ERIHMALQSEESTPKNGDWAVKLGINLFYSAILS
						RSPLYNKQMPYNSIIYNAFGCTSPSKSSPEEAKV
						RGKGFGRQKKLVMAGKWCGKVWMSNQVHPLL
						LHRDPDEEERNFNACMKSDEKVGRKSETSHKAQ
						TTYTNRKVGKKRRSMPESTKVKKMKFEATEFAD
1016	sp_185680_rszf	-0.851354	-0.814384	-0.873111	3523 (source: NCBI BLink).	TDPEDSVDDKPEVETRRSVRKSKTQCRSIKRVK
						MGNIFVKKPKITEVDRAILSLKTQRRKLGQYQQQ
						LEAVIEAEKQAARDLIREKRKDRALLALKKKKLADI
					vacuolar protein sorting-associated protein 20.2 (VPS20.2); INVOLVED IN: vesicle-mediated transport, N-terminal protein myristoylation; LOCATED IN: ESCRT III complex, plasma membrane; CONTAINS InterPro DOMAIN/s: Snf7 (InterPro:IPR005024); BEST Arabidopsis thaliana protein match is: SNF7 family protein (TAIR:AT5G63880.1); Has 1828 Blast hits to 1807 proteins in 295 species: Archae - 35; Bacteria - 99; Metazoa - 731; Fungi - 367; Plants - 251; Viruses - 5; Other Eukaryotes - 340 (source: NCBI BLink).	ELSSKQKAVFESLKAGNNTMKAIQSELNLDDVQK
1017	sp_186130_dyxo	-0.884379	#N/A	#N/A		LMDDTEEARAYQE
1018	sp_186271z9033	#N/A	-0.910123	-0.926227	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1019	sp_186300_kxyq	#N/A	-0.3332226	-0.3831134	Protein of unknown function, DUF593; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 18 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF593 (InterPro:IPR007656); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF593 (TAIR:AT4G30830.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MSLMATRTLSLATHGSNFLKET_QAQQELLQKLY NELDMERESSATAADEALSMILRLQGEKAAVGM EASQYKRIAEEKISHVQETLEVFQEIIYQKEMQIAS LEFQVQAYRFKLLSMGCEDLGAGELHFPDNMLS RKNVVSEESSTRSLRRLSSLPPNPLNGVLSRRTS NEGTTRSSTSSPIPWQESKSADPSVRKEHGDFH SYWDEIRKLDDRVKELDSLKVSNSPPSTEDTNDV PLTKKSAIMPHPSPENFPAKTNFVESLEKVASVQ ACESSACSTTAQDIYEVPICSYIDSSNIKKESQNV TLGSETRNGKLDLVWGEIWKSDVKEPLLSPESD NSSKTNDSKDKIDWEKKPLLSPGSSNESPKLKD RKTDKPLLFKHPTLGGIADFELQQFNRRLKQLED QRFVTRGHEMNSEKAEEEELRVLREIRDKIDLMQ ADITRWGIKKPLPADSRSPGCLKEVILCFWPS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1021	sp_187500_fcis	-0.853313	-0.887372	#N/A	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT4G00830.2); Has 31429 Blast hits to 23398 proteins in 2969 species: Archae - 812; Bacteria - 21040; Metazoa - 874; Fungi - 1027; Plants - 329; Viruses - 0; Other Eukaryotes - 7347 (source: NCBI BLink).	MPPRTKKSASTPAKKSGRLSRTAAKTEAKSDDGI NEVVRDEINDNVKVSELDNDVDSSADLNVEPQS QANGSFILTKEDEVKEDEYEKDERLDLEDNEPEY EPEEYGGEYDEKEVDQDDFQEDGDEGMEEVEE SDLAEEEEEEEEGDMAEEDIEEVPEELGVEEVDE HADADADEGREHQHHHDHGHDDEHAELFDAAE EREHHELVKERRKRKEFEVFVGGLDKDATEEHL KEVFSVVGEVTEVRLMMNPQTKKNKGFAFLRFS TIEQAKRAVSELKNPVVNGKQCGVTPSQDSDTLF LGNICKTWTKEKLKEKLKLYGIENIEDLTLVEDSN NEGKNRGFAFLEFSSRSEAMDAFKRLQKRDVLL GVERPAKVSFADSFIDPGDEIMAQVKTVFIDGLPA SWDEDRVRILLKNFGEIEKVELARNMPSARRKDF GFVSFDSHDAAVTCAKSINNAELGEGDHKAKVR ARLSRPLQRGKGKNPRVDHNKGGRVAGRRPRA PWNRPLPRPLPPRMRAIGSRPPIVVGRGLKRPV GVRDRRPPPPMPARARPLPPPPRSYERRAPGPP PPKSSLKRDYIRREELPPRSRAVAEYGSRLPPER RPSYRDEYAARSPVYPDVHRSSASRSATRRPYP EDNYPPRYERPPPSYREGRPRDYDSISGSKRPY SAMDDVPSRYTDPGHGMRHSRARVDYDVPSGS SQYADTYGDRMGRSTAGYSGSGSRSSVSAQDS HGLYNSRQGMSYNGSYSGNNDGVYSSSYSSEY ASRGSDVGGSSYSSLYSGRGMSGSGYMGGGS SGSYY
1022	sp_188550_huor	-0.891484	#N/A	#N/A	TIMING OF CAB EXPRESSION 1 (TOC1); CONTAINS InterPro DOMAIN/s: CheY-like (InterPro:IPR011006), Signal transduction response regulator, receiver domain (InterPro:IPR001789), CCT domain (InterPro:IPR010402); BEST Arabidopsis thaliana protein match is: pseudo-response regulator 3 (TAIR:AT5G60100.3); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MEIREVGPSNGKSTDQFIDRSKVRILLCDNDSKS SQEVFSLLCKCSYQVTSVKSARQVVDALNAERS EIDIILSEVDLPTKKGFKLLKYIMRDEQLQRIPVIM MSTQDEVSLVVKCLRLGAADFLVKPLRTNELLNL WTHMWRRRRMLGLAEKNIVSYDFDLAASDHSE NSTTLFSDDTDDKSRRSVNPEISFTMQQEEECN KVTSDAVPANSFDCRSTLPALGDRRTGFVPFSK KGEPFLKKSEPFPKRRELKIGECSAFFTYVKTGA VESNCQTVVDVEGNAAKPSRPDDKLDTWNGHVI NGNMTYENGDVWENNSHGEDFPSSTSVPDSTS MERSCTPPVTVEFMQNKSNDEGLSEVHMQPKN GKPHFDVSTLAAHAAAYPYFMSGVMNQFMMPS SAHANQKDLQNHGSSTMMPHYNHVSHSSSQVP GMAPFTYYPLGVCLPQGQTPANHQVPSFGNSFS SEAKLSKVNRREAALIKFRQKRKERCFDKKIRYV NRKKLAERRPRVRGQFVRKVNGVTVDLNGDPSA DLYEEDEDEPNSSPEDDTF

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1023	sp_189180_odsd	-0.892742	#N/A	#N/A	uncharacterized protein	MVVDLMTPTRQNQEQEVNDTDVAGMGQKTPPL MPPTLHTSSGYVRPSQSAGTSSGLNMGVFVRP GEWIEKTPICLSPKKMKYFEVPPGEPDEPWNPKI DLLRGESILTDDCKGGGCMGWRALKDLATPR
1024	sp_189400_twmw	-0.885639	-0.879935	-0.857472	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT5G47620.2); Has 54138 Blast hits to 30457 proteins in 1475 species: Archae - 38; Bacteria - 13880; Metazoa - 18730; Fungi - 4978; Plants - 7803; Viruses - 149; Other Eukaryotes - 8560 (source: NCBI BLink).	MEPGKLFIGGISWDTNEDRLREYFETFGEVVEAV IMKDRATGRARGFGFVIYADPAVAERVVRQKHMI DGRTVEAKKAVPRDDQHVLTRNHSSPIASPVPA RTKKIFVGGLPSTVTEGDFKRYFDQFGIITDVVVM YDHNTQRPRGFGFITFDSEDAVERALMKTFHELN GKMVEVKRAVPKELSPGPSRSPLGAGSGGLTRV NSFLNGFTQGYNPSPIGGHGVKIDERQSPMPFA RGGFPSFSPSSYGSALNYEPRMSPGFGETANLY PKLSFEQGLSALHNGNINRHDNFVEFDGVNRGS GSTIGSSSWSLWGNGNGNGNGNGNGSINDSPF SGNNASMNFRDIEADALKNTEFIWNTSAIPVQGG GRASVSNGLSLSYGREDNNLGSAMGLYGRDIRT RTQPVSSFPSLDVGFDAKFDDFLGGGGEQSLLL PSLGLKGSGSFEFGLGNSVSDVPLNNPAGLCW
1025	sp_189670_pjdr	-0.933019	-0.970866	-0.913293	PIF / Ping-Pong family of plant transposases; LOCATED IN: chloroplast; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Putative harbinger transposase-derived nuclease (InterPro:IPR006912); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G12010.1); Has 644 Blast hits to 642 proteins in 76 species: Archae - 0; Bacteria - 0; Metazoa - 379; Fungi - 49; Plants - 189; Viruses - 3; Other Eukaryotes - 24 (source: NCBI BLink).	MLFDQFGLLIVDVFSLLVLIKETEMDQGFLVMLSN LLHLQNHLDPSSSVFSSDATSLSSSSTSSVSSLL TSSSAAPLLFFTIASVLSFVAASRPSSSSSPTSSS RSPSRSRSSSPSDHSLSAFRALSTELIWSMDPPV RDAQWRSSYGLSYPVFTTVVDKLKPYIAQCNLSL PPDYAVAMVLSRLSHGLSAKTLAKRHSLDPYLVS KITNMVTRLLSTKLYPEFIKIPAGRRRLHETTQAFE ELTSLPNLCGAIDSTPVKLRRLPTDVSDPSSYHC RYGYPAIQLQVVADHKKIFWDVCVKAPGSVDDA THFRDSVLYNKLMSGDVVWDKVVNVRGHPVRP YIVGDWCFPLLSFLLTPFSGTGRGTPAQNTFNAA LMKARLASVEAVGLLKGRWRILQDLNVGLHHAP QTIVACCVLHNLCQIAREPEPELFKEPEENGSLP RVLESERSFYYYGENLRQTLADDLQQRLSSR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1026	sp_189760_jspe	#N/A	-0.87:846	#N/A	LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (LOS1); FUNCTIONS IN: translation factor activity, nucleic acid binding, copper ion binding, translation elongation factor activity; INVOLVED IN: response to cold; LOCATED IN: cytosol, nucleolus, chloroplast, plasma membrane, membrane; EXPRESSED IN: 29 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein synthesis factor, GTP-binding (InterPro:IPR000795), Translation elongation factor EFG/EF2, domain IV (InterPro:IPR005517), Translation elongation factor EFG/EF2, C-terminal (InterPro:IPR000640), Translation elongation factor EFTu/EF1A, domain 2 (InterPro:IPR004161), Ribosomal protein S5 domain 2-type fold (InterPro:IPR020568), Ribosomal protein S5 domain 2-type fold, subgroup (InterPro:IPR014721), Elongation factor G/III/V (InterPro:IPR009022), Translation elongation/initiation factor/Ribosomal, beta-barrel (InterPro:IPR009000); BEST Arabidopsis thaliana protein match is: Ribosomal protein S5/Elongation factor G/III/V family protein (TAIR:AT3G12915.1); Has 64297 Blast hits to 57476 proteins in 4618 species: Archae - 1324; Bacteria - 41101; Metazoa - 3172; Fungi - 1934; Plants - 1269; Viruses - 1; Other Eukaryotes - 15496 (source: NCBI BLink).	MVKFTADELRRIMDCKHNIRNMSVIAHVDHGKST LTDSLVAAAGIIAQETAGDVRMTDTRADEAERGIT IKSTGISLYYEMTDEALAAFKGERAGNDYLINLIDS PGHVDFSSEVTAALRITDGALVVVDCIEGVCVQT ETVLRQALGERIRPVLTVNKMDRCFLELQVNGEE AYTTFQKVIENANVIMATYEDPLLGDCQVYPDKG TVAFSAGLHGWAFTLSNFAKMYASKFGVDESKM MERLWGENFFDPATKKWTSKNTGTATCKRGFV QFCYEPIKVVIATCMNDQKDKLWPMLAKLGVQLK SEEKELMGKALMKRVMQTWLPASSALLEMMIYH LPSPATAQRYRVENLYEGPMDDVYATAIRNCDP EGPLMLYVSKMIPAPDKGRFFAFGRVFAGKVAT GMKVRIMGPNYVPGEKKDLYVKNVQRTVIWMGK RQETVEDVPCGNTVAMVGLDQFITKNATLTNEKE SDAHPIRAMKFSVSPVVRVAVQCKVASDLPKLVE GLKRLAKSDPMVQCSITESGEHIIAGAGELHLEIC LKDLQDDFMGGAEITKSDPVVSFRETVLEKSSRV VMSKSPNKHNRLYMEARPMEDGLAEAIDDGRIG PRDDPKIRSKILAEEFGWDKDLAKKIWCFGPETT GPNMVVDMCKGVQYLNEIKDSVVAGFQWASKE GALAEENMRGICFEVCDVVLHADAIHRGGGQVIP TARRVIYASQLTAKPRLLEPVYLVEIQAPENALGG IYSVLNQKRGHVFEEMQRPGTPLYNIKAYLPVVE SFGFSSTLRAATSGQAFPQCVFDHWEMMMQDP LDPTSQAAALVQNIRKRKGLKEQITPLSDFEDKL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1027	sp_190700_syta	0.827306	0.861127	0.914224	DNA polymerase V family; FUNCTIONS IN: DNA-directed DNA polymerase activity, DNA binding; INVOLVED IN: DNA replication, transcription; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: DNA polymerase V (InterPro:IPR007015); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MGSRKGGPSSVEEEIVVEGETSNEKLGLDKTKG KKGKKGKDNGPPSEGGSDGPSTPASSSAKPME RRKKRKLLDKVRHSSATESKEIKPAQETPVWKFE DPGTASSSNGQPVLHMEVFTELASADASKREAA AERLAVELQEIQRAYKQGGNEDSVQGGLQLEAA KDDGLDNCAPPVRYAVRRLIRGVSSSRECARQG FALGLTIVVSTNSKIRVDSVLKLIVDSLEVTSSMKG QEARDCLLGRLFAYGAIARSGRLKEECASDINSP YIKEFTSHLMSVASKKRYLQEPTVSVLVNLMDML PIEAVGNHIVYAPGLREWFAAAPTEGNPDALLLA LKIRQKTATDSVVFDDLLPRPFSPSRFFTVDHLTS IANCLKESTFCQPRVHGIWPLLVSLLLPDTVLQDA NSSSSLSSTKKNKKSRRSSSSDEDIAESLRCFCE VILEGSLLTSSHDRKHLAFDILLLMLPRLPAFHIPIV LSHKFVQGLMDILSTKDSWLYKVAEHFLKQLLD WVGSDDIKRVFVVVALQKHSGGRFDCITRTKTIK DLMAGFTTESGCMLFMQKLTSMFVNEEHASDEP SDKNLATDNMSEIGSVEDKESADSNSDLMKSWI VESLPTLLKYLTLDPEAKFRVQKEVMKFLAVQGL FSASLGTEVTSFELLQEKFEWPKTPISSALCKMCI DQLQLLLSNAQKAEGPHALPNSLETNDLGCYFM RFLGTLCTIPSVSLVRPLSNEDEKTFKKLQALEST LSREERNRGLSADANKLHALRYLLIQLLLQLLQRP GEFTEAASELIICCRKAFSIHDPPESSEEDEAEDG ESPELMDVLVDTLLSLLPQSSAPLRSAVEQVFKY FCNDVTDDGLLRMLRVIKKDLKPSRRKDADSED GSDADDDDLLDLEEAEESDEADAVETAVMTDDT EATVTADASDEEFPDASDSDDDSDEDMDDEAMF RMDSYLAKIFQERKNQAGGDTAHAQLILFKLRVL SLLEIYLHENPGSSRVLTVYSNLVQAFVNPSTTE GSEQLAQRIWGILQKKIFKAKDYPKSEDVQLVTL EGLLEKSLRLASRPFRKKKTSGSVQQKKKSFSW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1028	sp_192520_rhen	#N/A	-0.896977	-0.912252	RNA helicase family protein; FUNCTIONS IN: helicase activity, nucleic acid binding, ATP-dependent helicase activity, ATP binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Helicase-associated domain (InterPro:IPR007502), DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), Domain of unknown function DUF1605 (InterPro:IPR011709), DNA/RNA helicase, ATP-dependent, DEAH-box type, conserved site (InterPro:IPR002464), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEA(D/H)-box RNA helicase family protein (TAIR:AT5G04895.1); Has 15667 Blast hits to 10761 proteins in 1709 species: Archae - 0; Bacteria - 5024; Metazoa - 4470; Fungi - 1848; Plants - 1329; Viruses - 777; Other Eukaryotes - 2219 (source: NCBI BLink).	MSMRCFPFPTTSVPTTFSLLKHFTSSFPFPISKPP LLHLSPRLFLSSMSSRPNYQGGRRGGGGGGGR NRGGGNRGGGSGGRGGRGGGGGGGEQRWW DPVWRAQRLAQKAAEYEVLDENEWWNKMEGF YKGGDQELIIKHLFRREDQQILFDMAQQLGLHFH AYNKGKTLVASKVPLPNYRADLDERHGSTQKEIR MSTDIERRVGSLLDTSKGTSSTENSIGESSQSSK NSSSTISSVKPVPILESHVAKEKISAELKHQEEKL KASASVQAMQSFRAKLPAFNVRSKFLEAVAKNQ VLVVSGETGCGKTTQLPQFILEEEISLLRGADCSII CTQPRRISAISVAARISSERGENLGETVGYQIRLE TKRSARTRLLFCTTGVLLRRLVQDPNLTGVSHLL VDEIHERGMYEDFLLIILKDLLPRRPDLRLILMSATI NAELFSQYFGNAPSIHIPGLTFPVAEFFLEDVLEK TRYSIKQESDSFQGSSRRRVKQLDSKKDPVTELF EDVDIDSHYRSYTINTRKSLEAWSSLQLDLSLVQ STIEHICRHEGEGAILVFLTGWDEISKLLDSIKANH FLGDTRKFLVLPLHGSMPTVNQREIFDKPPASTR KIVLATNIAESSITIDDVVYVIDCGKAKETSYDALN KLACLLPSWISKASAHQRRGRAGRVQPGVCYRL YPKLVYDAMLQYQLPEILRTPLQELCLNIKSLQLG TVGSFLAKALQPPDSLSVQNAIELLKTIGALDERE DLTPLGRHLCTLPLEPNIGKMLLMGSVFQCLNPA LTIAAALAYRNPFVLPINRKEEADAAKKSFAGDSC SDHIALLKAFEGWKEAKSSRTEKSFCWENFLSP VTMQMMDDMRKQFVDLLSDIGFVDKSKGPNAYN QYSHDMEMVCAILCAGLYPNVVQCKRRGKRTAF YTKEVGQVDIHPASVNAGVHLFPLPYMVYSEKVK TTSIYIRDSTNISDYALLLFGGSLLPSKSGEGIEML GGYLHFSASKSVLELMKKLRLELDKLLSRKIEEPR LDITAEGKSVVAAVVELLHS
1029	sp_193100_dzum	-0.876358	-0.883555	-0.923894	unknown protein; Has 1784 Blast hits to 634 proteins in 116 species: Archae - 0; Bacteria - 0; Metazoa - 1013; Fungi - 200; Plants - 288; Viruses - 0; Other Eukaryotes - 283 (source: NCBI BLink).	MVKGIDELTEDERKALRRRKFAPLPPKPSSSCSS SLPRLAHPGGPVTTNKAVALAKFLERKLQEPGGL DTLKPELIELAVKNAKETVNASGTSKHGKIIRHVT SFDDNEDYSSKDWPTYEPKIKLKKEKKDKKKKKK KKKKNKKNKKVDDQ
1030	sp_193160_ktmf	-0.977898	-0.991893	-0.95987	uncharacterized protein	MIQPLYLVTRGDEIIHLVLYIGEDGLHILQLLIHRSF GELDDDEETKHHFIDLKLSKGRFRL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1031	sp_193390_qjyj	#N/A	-0.804297	-0.894435	SNF7 family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: protein transport; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Snf7 (InterPro:IPR005024); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MAEELKVVEFIREQVPDWDDDVKATARFKAFSG QRSDWESQFFFWRDLIIKIAREFDLLFLSPSQMK NCWFNRGGLSPLCLNDSLLEMYKAGNVLRVTDF GDPTSGRLSQLLKKLIHFSIVFRPSTSNDILNDRLI LRVMLEEKSADVIKALSETHWTSSCIITMRKFESL FREHNEAYAVLCYLSGYGKARYISVKRKEELIEG VKVSLSSAAAPILPSTDSAILQLVWTLEELQKQLL VIDEQYKRLRISALAALKSGDKKLALRNARQLKLA SESREKLTSLFNRVEEVLRAITDAESAKKVSEAIQ LGALALKENRTDVEKVQMCLQELDEFRDSQEIIH NALESTSYTAMEDEDIEEELKRLDSQIRDENLMV SALKSKDDAIRNAGASVTPDSLCAAIVDLKLESET QETTQGPGNRTSLELEAA
1032	sp_193890_zsia	-0.8683883	#N/A	#N/A	uncharacterized protein	MFNMTVQTQNT
1033	sp_194030_anpy	#N/A	-0.855305	#N/A	uncharacterized protein	MCNLPFVYSLFTPQTHLFLHPLQKDDIAIKAATEM VRTKPILVQLGPTVALSKSNALQWCLKKVALEEN TEEASEENEENDQAALEASEENDQAALENTAAP KK
1034	sp_194100_fstz	-0.881038	-0.9461178	-0.9193883	uncharacterized protein	MKNVAKCDTWCELQNPVNHRVFERKLRPKPSG RGHACLGVTHRVSPNPRTAGRSRRMASHPSPG VDGLNKEPPVTNCRSDRW
1035	sp_194830_isew	#N/A	#N/A	-0.8803366	uncharacterized protein	MAPLLQDLRAVIREQVIRRLGAKESVAEKDPSPG VAEKDPSPGVAEKDPSPASSSHRQISDNAYINVD GEPMAVDKSVNKVLSELDVIVLEDEIDNVQHRVE QPLGGEQNVFVNYQSLPVEQPVFEFLHVPAVEA VQNEPMEAEQVQQGP
1036	sp_195680_enhj	-0.9823011	-0.923439	-0.8823027	uncharacterized protein	MRNCLERMYLNITAHERIWNSKNTTSMLKVVICIP YSRLGCQLHFPESSFVRIVVYGMP
1037	sp_195880_ggoo	-0.974449	-0.919651	-0.95437	uncharacterized protein	WPTRHLTRLDGCSRSPYSASPLSACSYCCSGAA AQCCCCLSTRSATGRRSASGSPVSVE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1038	sp_196220_mdwg	-0.886043	#N/A	-0.6453857	DEAD/DEAH box RNA helicase family protein ; FUNCTIONS IN: helicase activity, ATP-dependent helicase activity, nucleic acid binding, ATP binding; INVOLVED IN: response to cadmium ion; LOCATED IN: nucleolus; EXPRESSED IN: guard cell, root, cultured cell; CONTAINS InterPro DOMAIN/s: DNA/RNA helicase, DEAD/DEAH box type, N-terminal (InterPro:IPR011545), RNA helicase, DEAD-box type, Q motif (InterPro:IPR014014), DEAD-like helicase, N-terminal (InterPro:IPR014001), DNA/RNA helicase, C-terminal (InterPro:IPR001650), Helicase, superfamily 1/2, ATP-binding domain (InterPro:IPR014021); BEST Arabidopsis thaliana protein match is: DEAD/DEAH box RNA helicase family protein (TAIR:AT5G11200.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MGETRENDAYEEELLDYDEEEEKVPDPAVTKTN GEAAKKGYSGIYSSGFRDFLLKPELLRSIVDCGFE HPSEVQHECIPQAILGMDVICQAKSGMGKTAVFV LSTLQQIDPVAGQVAALVLCHTRELAYQICHEFE RFSTYLPGVKVAVFYGGVNIKIHKDLLKNECPHIV VGTPGRILALARDKDLSLKNVRHFILDECDKMLES LDMRRDVQEIFKMTPHDKQVMMFSATLSKAIRPV CKKFMQDPMEIYVDDEAKLTLHGLVQHYIKLSEL EKNRKLNDLLDALDFNQVVIFVKSVSRAAELNKLL VECNFPSICIHSGMSQEERLARYKGFKDGKQRIL VATDLVGRGIDIERVNIVINYDMPDSADTYLHRVG RAGRFGTKGLAITFVSSASDSDVLNQVQERFEVD IKELPEQIDTSTYMPS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1039	sp_196330_cuqo	#N/A	-0.853659	#N/A	mitogen-activated protein kinase kinase kinase 7 (MAPKKK7); FUNCTIONS IN: protein serine/threonine kinase activity, binding, protein kinase activity, kinase activity, ATP binding; INVOLVED IN: plasma membrane organization, pollen development; LOCATED IN: cytosol, plasma membrane; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: ovule developmental stages; CONTAINS InterPro DOMAIN/s: Protein kinase, ATP binding site (InterPro:IPR017441), Serine/threonine-protein kinase domain (InterPro:IPR002290), Armadillo-like helical (InterPro:IPR011989), Armadillo (InterPro:IPR000225), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), Protein kinase, catalytic domain (InterPro:IPR000719), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: mitogen-activated protein kinase kinase kinase 6 (TAIR:AT3G07980.1); Has 137235 Blast hits to 134768 proteins in 5401 species: Archae - 190; Bacteria - 15422; Metazoa - 52001; Fungi - 13364; Plants - 33616; Viruses - 670; Other Eukaryotes - 21972 (source: NCBI BLink).	MSRQMSSAREFRKSKTLDNKYMLGDEIGKGAYGR VYKGLDLENGDFVAIKQVSLENIAQEDLNIIMQEID LLKNLNHKNIVKYLGSLKTRSHLHILEYVENGSLA NIIKPNKFGPFPEQLVAHYIAQVVEGLVYLHEQGV IHRDIKGANILTTKEGLVKLADFGVATKLTEADVN THSVVGTPYWMAPEVIEMSGVCAASDIWSVGCT VIELLTCVPPYYDLQPMPALFRIVQDEHPPIPDGL SADMADFLRLCFQKDARLRPDAKTLLSHPWILNY RRTLQPSLPASDAIRNTQEGLMDTELSNGNENG VAESHSVDEVARRETLLTDADTSTADDNQGLSE EVSGFCDDLSADQVPTLAIHEKSVAKSTSMISARI QGAGSEQISEQGHTIDPDEVLTNGEAASPKSKRK DGVSGKVDGKGSAVHKMNSLFSFGTRNQDENL QKVAKKGISTGNELSKFSDTPGDASLEDLFHPLD KNTEDRLADTSASTSTSHLNPSSTTLESGKNDLA TKLRAAIAQKQMENEMGQANGGNLLRVMIRALK EDVIDIDSSLVFDDKMPGENLFPLQAVEFSRLIAS LRPEEPEDVIVNACLKLIAFFNQRSEQKVVFVKQ HGLLPLMELLEVPRPRVICSVLQVINQIIKDNTDFQ ENACHVGLIPIVMGFAVPDRPREVRMEAAYFLQQ LCQSSSLTLQMFIACRGIPVLVGFLEADYAKYRE MVHLAVDGMMWQVFKLQKSTPRNDFCRIAAKNGI LLRLINTLHSLNEATRLASISGVSGLTSDGSVQRP RSGQLDPSHPVFSPNEATHSGSDPVDLRVRHSA VEHSSSTVDADKPQSSHAIVESSASSRSTDAADK FVNVVKEATRAEIDLRQHRVAGSSNRTSTDRPPK VEGGSNGLPYAVSSQQDQVRPLLSLLEKEPPSR HFSGQLDLRQLAGMERHESILPLLHASNDKKGN GELDFLMAEFAEVSGRGRENGNSETPRISNKTA VKKLISVASNDGAASTSGLASQTASGVLSGSGVL NARPGSATSSGLLSHMVSAMNADVSREYLEKVA ELLLEFAQADTTVKSYMCSQSLLTRLFHMFNRVE	
1040	sp_196400_wumz	#N/A	-0.856272	#N/A	RHOMBOID-like 2 (RBL2); CONTAINS InterPro DOMAIN/s: Peptidase S54, rhomboid (InterPro:IPR002610); BEST Arabidopsis thaliana protein match is: RHOMBOID-like protein 6 (TAIR:AT1G12750.3); Has 6856 Blast hits to 6855 proteins in 1947 species: Archae - 152; Bacteria - 4411; Metazoa - 521; Fungi - 179; Plants - 374; Viruses - 0; Other Eukaryotes - 1219 (source: NCBI BLink).	MGSRDLEKNGDVNSSYNRYNSYNTQQPNHPVA PSGYHIESSARQWTSWLVPMIVVANVAMFIVVMF INNCPKNNFGFDGCVARFLGRFSFQPLRENPLLG PSANTLMNLGALQWDKVVHQHQGWRLITCIWLH AGVVHLLVNMLSLVFIGIRLEQQFGFVRIGVLYVIS GLGGSVMSALFLQKSISVGASGALFGLLGAMLAE LLTNWTIYANKAAALFTLLIMIAINLAVGILPHVDNF AHIGGFISGFLLGFVLLLRPQYGWLERHQYPEHA RVKSKFMPYQMVLFFIGAALLITGLTIALVMLFQG KNGNDHCRWCRYLSCVPTSKWQCGN	
1041	sp_197361z10483	#N/A	-0.833561	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1042	sp_197390_sacn	-0.971451	-0.951764	-0.923082	small ubiquitin-like modifier 2 (SUMO2); CONTAINS InterPro DOMAIN/s: Ubiquitin (InterPro:IPR000626), Ubiquitin supergroup (InterPro:IPR019955); BEST Arabidopsis thaliana protein match is: small ubiquitin-like modifier 1 (TAIR:AT4G26840.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSGTQEEDKKPGGDQAAHINLKVKGQDGNEVFF RIKRSTQLKKLMNAYCDRQSVELNSIAFLFDGRR LRAEQTPDELEMEDGDEIDAMLHQTGGAMI
1043	sp_197640_frpi	-0.934283	-0.940742	-0.934434	CW7; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF2045 (InterPro:IPR019141); Has 145 Blast hits to 144 proteins in 52 species: Archae - 0; Bacteria - 0; Metazoa - 91; Fungi - 0; Plants - 50; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	MKGPGGRGEVEVAVSGVEDQSEHSSGPSSPGS SKGAFGLATIVRRAASVASVAAKQAYAAASNNAD DDMLPLNCSLVSVSLPWEHIAHDLLLKVGPPVSM
1044	sp_198000_xzyn	#N/A	-0.874341	-0.90317	profilin 5 (PRF5); CONTAINS InterPro DOMAIN/s: Profilin/allergen (InterPro:IPR002097), Profilin, plant (InterPro:IPR005455); BEST Arabidopsis thaliana protein match is: profilin 4 (TAR:AT4G29340.1); Has 915 Blast hits to 914 proteins in 244 species: Archae - 0; Bacteria - 2; Metazoa - 138; Fungi - 121; Plants - 583; Viruses - 0; Other Eukaryotes - 71 (source: NCBI BLink).	MSWQTYVDDHLMCEIEGTTNHLAAAAILGTDGS VWAQSASFPQFKPDEIAAVVKDFDEPGTLAPTGL HLGGVKYMVIQGEPGAVIRGKKGSGGICVKKTA QALVIGLYDEPVTPGQCNMIVERLGDYLIDQGL
1045	sp_198640_nawc	#N/A	-0.926733	-0.894179	PAR1 protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: PAR1 (InterPro:IPR009489); BEST Arabidopsis thaliana protein match is: PAR1 protein (TAIR:AT3G54040.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MASHFNMSALVTLSILLSSLLIQGTLGEVTCEQLN KEACAFAVSSIGKRCVLEKYVRRSGLEAYKCTAS SIEADKLKDLVETNHCIEACGLYRETFGISSDSLL ESRFTQKLCSPECYQRCPNVVDLYFNLAAGEGV FLPKLCETQSENPRRGLSDIKSSGIVAPGPVSGN DDSKLFGITSIGSPLPSMELAAPAPAPAPAMPPL

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1046	sp_198840_rzcc	-0.852312	#N/A	-0.834854	ARM repeat superfamily protein; FUNCTIONS IN: protein transporter activity, binding; INVOLVED IN: intracellular protein transport, protein import into nucleus, docking; LOCATED IN: nucleus, nuclear pore, cytoplasm; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Importin-beta, N-terminal (InterPro:IPR001494), Armadillo-like helical (InterPro:IPR011989), Armadillo-type fold (InterPro:IPR016024); BEST Arabidopsis thaliana protein match is: ARM repeat superfamily protein (TAIR:AT3G59020.2); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MTNAIDLDHDQQWLINCLNATLDTNQQVRSFAEA SLNQATLQHGFGSALAKVAANRELPMGLRQLAA VLLKQFIKKHWDEDDEDNFEHPVVACDEKEVIRR LLLLCLDDSHRKICTAVSMAVASIAHYDWPENWP ELLPFLMKLINDQTNMNGVLGALRCLALLSGDIDD KAVPSLVPILFPCLLSIVSAPQIYEKSMRTKALSIV FYCTSMLGTMSGAYKAETSYLMSPMLQPWMEQ FSAVLSHPVQPEDPEDWSIRMEVLKCLNQFVQN FPSATEKEFTSIIGPLWQTFVTTVKVYEKSSVEGA QDSFDGRYDSDGAERSLESFVIQLFEFLLTIVGSS RFVKVVADNLQELVYYTFGFMQITEHQMHAWSL DANQYISDEDDNMYNCRVSGVLLLEEIISSCGED GLYTLVGAARRRFIESQQQKADGFEFWWFRIREA VIYAMASLSDQFQEAEVCGPDGFDLKKFVEQMIM EDVGAGVHECPYLHARVFSSVPKFSSMLSPGVV EHFLFASLRAIGMDVPPVKVGSCRVLSQLLSEA NKEALQPQILQLFSSLIDLLNQASEESLHLVLETL QAAIMSGHEASLSVEPVISPIILNTWVSHVSDPFIS IDAIEVLEAIKNAPGCIQPLVSRVLPYIGPILSQPKQ QADGLVAGSLDLVTMLLKNAPGPVVKALYDVCF NPIIQIVLQSDDHSEMQNATECLAAFVCGGKSELL AWGGDSGFTMRSLLDALS
1047	sp_199140_gfod	#N/A	#N/A	-0.888348	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Craniofacial development protein 1/Bucentaur (InterPro:IPR011421); Has 333 Blast hits to 324 proteins in 149 species: Archae - 0; Bacteria - 18; Metazoa - 117; Fungi - 96; Plants - 49; Viruses - 0; Other Eukaryotes - 53 (source: NCBI BLink).	MEGIPIESKKETASVSASQVQSEDHEIKARVDAV WQQMNKGVSVKTMNSIVNKSSSTGTKTMQTDK ASNNWMSLLGLGSKKTKPTVASTKGKNSSPAQN GTSDEARKLAAAALAAARDAATASASAMNRGKIE VTEVRDFAGEEIEVKKLIDPSSKEAAERAGPSSA VDSVLEQIKKKPKLNLLDKTKKDWGEFKGENKDY EEELDAYKKSSNQYLDKVSFLSRADYREFERER DVRLAQQAKRKPDT
1048	sp_199170_awmz	#N/A	-0.909306	#N/A	uncharacterized protein	MESGDEARHRRLLRRFILIQRRNRARNNNPTHV MVTDSRPRHLPVIWPQHPMPSMMNAVQRVGMD NGDEAKHRRLLRRFILIQRRNRARNNNPTHVRVT DSRPRHLPVIWPQHPVASMMNAVERAHRSLTSR GETSAPNNNVIGSGTPSRRRRRSENRESTSSLR RRLNSSLPPDGSMTPTTETDIQNTRSPQPNLNRV GSNISRNTLEGSASHALHHRKYTRPWTFDTTTG VETPLLDNATGSNTTVARHLLDNASENDTRVERT ARKRNRNEMETDKETTTTANREN
1049	sp_199381z10789	-0.930365	#N/A	#N/A	uncharacterized protein	#N/A

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1050	sp_199410_irgg	-0.8307776	#N/A	#N/A	ribonuclease 2 (RNS2); FUNCTIONS IN: ribonuclease T2 activity, endoribonuclease activity, RNA binding; INVOLVED IN: aging; LOCATED IN: intracellular, vacuole; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Ribonuclease T2 (InterPro:IPR001568), Ribonuclease T2, active site (InterPro:IPR018188); BEST Arabidopsis thaliana protein match is: Ribonuclease T2 family protein (TAIR:AT1G14220.1); Has 2318 Blast hits to 2317 proteins in 471 species: Archae - 0; Bacteria - 380; Metazoa - 295; Fungi - 231; Plants - 1299; Viruses - 5; Other Eukaryotes - 108 (source: NCBI BLink).	MTMALSISIFALLLITFTVAVASTADVDSSLLLSSN QREFDYFQLALQWPGTYCRTTHKCCQKNGCCR NSGAPQHFTIHGLWPDYNDGTWPSCCTKSNFD EKEISTLRKDLDEYWPSLSCGSPSSCHGHKGSF WGHEVQWLPLTQYELDYICFPNMFLTVLLFCFLV PGSHDLRISGVKKHGTCSSPVVKDEYNYFRTGLT VYFKYNITEALSDAGYVPSNSEKYPLGGIISAIENA FHISPTVVCSGDAVKELNLCFYKDFKPRDCATRS RMDNSKTSCPKFVSLPTYDSSRLGIVGALESFVI
1051	sp_199570_huue	-0.8825551	#N/A	#N/A	STELAR K+ outward rectifier (SKOR); FUNCTIONS IN: outward rectifier potassium channel activity, cyclic nucleotide binding; INVOLVED IN: potassium ion transport, response to abscisic acid stimulus; LOCATED IN: membrane; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: L mature pollen stage, M germinated pollen stage, 4 anthesis; CONTAINS InterPro DOMAIN/s: Cyclic nucleotide-binding (InterPro:IPR000595), Potassium channel, voltage-dependent, EAG/ELK/ERG (InterPro:IPR003938), Protein of unknown function DUF3354 (InterPro:IPR021789), Ankyrin repeat-containing domain (InterPro:IPR020683), Ion transport (InterPro:IPR005821), Cyclic nucleotide-binding-like (InterPro:IPR018490), RmlC-like jelly roll fold (InterPro:IPR014710), Ankyrin repeat (InterPro:IPR002110); BEST Arabidopsis thaliana protein match is: gated outwardly-rectifying K+ channel (TAIR:AT5G37500.1); Has 82938 Blast hits to 34972 proteins in 1815 species: Archae - 147; Bacteria - 10317; Metazoa - 39460; Fungi - 7131; Plants - 3765; Viruses - 1138; Other Eukaryotes - 20980 (source: NCBI BLink)	EQGNVVDQLYLVCHGKLVTKVHEEFFLPGEVILE QGNVVDQLYLVCHGKLEEVSIGQDGSEQLISRLE PDSTFGQNSIFCNIPQSCTVRVVDLCRLLRIDKQS LSNIDIYFFDGKKIFDNLLKGNDGKFNLTEVESDI ASHITQQESELALKVNNAANHGDLYQLKSLLKAG ADLNKTDYNERSPLHIAASKGHENIVLFLIQAGAL VNQSDNFGNTPLLEAAKNGHDQIAAVLYREGARI DIKDVGTYLCTVVLSKESKDVDFLKRLLNNGVDP NSRDYDGRTPLHVACSHGLYLMAKLLVEAGALVL LKDRWDNTPLDEARLCGHNNLIKLLEAAKSAQLS QPAGQIEQISDRKLHKKCTVFPVHPWDTQEKRIP GIVLWIPRTIDELVKTASEKLNLPTATCVLSEDGG KILDIEMIDNGQKLYLITETD
1052	sp_200130_aejh	-0.910486	-0.883359	-0.850487	uncharacterized protein	MIIDAEKSGLR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1053	sp_200840_pngt	-0.854446	-0.819375	-0.858084	Protein of unknown function (DUF3537); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF3537 (InterPro:IPR021924); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF3537) (TAIR:AT3G20300.1); Has 140 Blast hits to 140 proteins in 15 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 138; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	MAVSLPPSHYSPLEEQWYNPDSAQTETLILPLKH SLRRLKTFLRFLGFIHDSRFSLSLSSFSFLLFGIAA PSLIIVFFYCGSGSASASACRKYHIKGFELETLFF QATAAAVSLLCFSLNFRKYGLRVFLFVDNYQQNV DRFQFEYIPKIDNFFRSLAVWVLPCVILKTAREIVR VIYVHNDSLVQSMMVVFVLLVSWTYTATVYLTGS ALFSLVCNLQVIHFEKYKNLLDRNLDVLVYIEEHS RLTFNLSKISHRFRIFLILEFLIVTASQIVTLLETTGN NGTINFINAADFAISSIVQVVGIIICLHAAAKISHRA QSLASFASRWHALLTCRSNETLGASENGGNLNV TGSAVAAMNYPESDMDSVDYILPPINEEMNSNVS SYLKRQAFVMYMQSNPGGVTIFGWTVNRAFMD TIFCVELTVVTFVLGKTLTFTTNNT
1054	sp_201140_jppd	-0.83288	-0.836245	-0.856635	AT-HSFB2B; FUNCTIONS IN: transcription repressor activity, sequence-specific DNA binding transcription factor activity; INVOLVED IN: response to chitin; LOCATED IN: nucleus; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: Winged helix-turn-helix transcription repressor DNA-binding (InterPro:IPR011991), Heat shock factor (HSF)-type, DNA-binding (InterPro:IPR000232); BEST Arabidopsis thaliana protein match is: heat shock transcription factor B2A (TAIR:AT5G62020.1); Has 32476 Blast hits to 10764 proteins in 847 species: Archae - 36; Bacteria - 10454; Metazoa - 9035; Fungi - 1651; Plants - 6330; Viruses - 666; Other Eukaryotes - 4304 (source: NCBI BLink).	MIPVPLDQNGDSTGCGGTGSGNASDLQQRTLPT PFLTKTYLLVDDPVVDDLISWNDDGTSFIVWRPA EFARDLLPKYFKHNNFSSFVRQLNTYGFRKVIPD RWEFANDCFKRGEKALLRDIQRRKISPAPAINSP TAVAACPTAIPAVVTVAAPPPVTVRAVSPANSGD EQVHSSNSSPAPGVVTVQRVASCASTTEILEENE RLRSENFTLSQELNHLRGLCNNILAMMSNYATGQ PETSGLQEGRPLDLLPTKPVTVAEADPCGASSG GATASGGSKGEEKMEEEEMEGCPCPRLFGVSIG AKRVRREDDEENRASVERSRSREGSDPIQTQQE QQCSGE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1055	sp_201480_aomj	#N/A	-0.908171	-0.926974	cytochrome P450, family 90, subfamily D, polypeptide 1 (CYP90D1); FUNCTIONS IN: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NADH or NADPH as one donor, and incorporation of one atom of oxygen, oxygen binding; INVOLVED IN: stamen development, petal development, leaf development, brassinosteroid biosynthetic process; LOCATED IN: endomembrane system; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Cytochrome P450 (InterPro:IPR001128), Cytochrome P450, E-class, group I (InterPro:IPR002401), Cytochrome P450, conserved site (InterPro:IPR017972); BEST Arabidopsis thaliana protein match is: Cytochrome P450 superfamily protein (TAIR:AT4G36380.1); Has 27722 Blast hits to 27661 proteins in 1582 species: Archae - 80; Bacteria - 4717; Metazoa - 10119; Fungi - 4209; Plants - 7293; Viruses - 6; Other Eukaryotes - 1298 (source: NCBI BLink).	MDILCSWTLISSLVIAISISLSYIIFHLYRNTKLKNFP KKLTWTLPLGTLGWPLLGETMEFISCAYSDHPES FMDRRRRKYGNVFKSHIFGKQTIISTDAEVNRFIL QSDAKLFVPCYPKSLTELMGKSSILLINGSLQRRI HGLIASFFKSSHLKAQVTRDMQCYVQQAMDTWV DNIHSPILIQNQAKHIAFQVLVKALISLDPGKEMHF LKKQFQEFVSGLMSLPVNLPGTQLYRSLQAKKE MVKLMKKIIVTKRNKMIQGADPNDVLDVLLSDQS GQLTDDLMSDNMIDMMIPGEDSVPLLITLAIKYLS DSPLALQKFTEENVKVNKLKARLGEPMYWNDYL SLPFTQMVRVLFRDEILSKN
1056	sp_201820_wssd	#N/A	#N/A	-0.869626	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G03460.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MDEAKQLMAQRQHQQLLMQQQQQQQQQQQQL YLLQHFQQQQQQQQQQQQQHQQQQQQQQQH HHQQQQQNQQQNQQQNQPQNQLHQHQQQRQ QQQAAMARFPSNIDAHLRPLGGANRPLGQPQN QNPNPNSNSIHPLHQQQNPSSSPQSQHQQLLQ QQLQQQKGMNRVLSQAQQQQAVPTLAGNQAEL QMAYQDAWRVCHPDFKRPFASLEDACERLLPYH VVADYEAEEDDRILDSDTTGQMFSRSQQWDHNI AAKVAEFTSTFEKQALAFNIITRKRALGEFRSEER LMIEQALLQEEKRAIMEVRAEIESRAGREAHEAK MRMAAMQAEQARAESQAHAELMARAPIRVNAL GSQGNDASMNHDMTDQEQGVIPDEMINGWGNN AQRVEREPSDDFLNDEEAENGDTGMQEEWRES GEFDLNSR
1057	sp_202750_xnjg	-0.958414	-0.99457	-0.956496	uncharacterized protein	MGLDQE
1058	sp_202791z11248	-0.879931	-0.868952	#N/A	uncharacterized protein	#N/A
1059	sp_202860_owok	#N/A	-0.877879	#N/A	uncharacterized protein	MAKKAAASWEGKWVASEKKLVDLAAVVKSKDE QLEGKDDQIADASKELERVKGELASTVEEMQGL RILYDALQEEFKDCEALAVWRTRAQMMFSCLKG EVSLWPCQKEVDDYLANGGTLEELSVPVVDADE LAMEADTAGTNGADADVAPLVEDVSSVDREGEA SMEQGEGDVPVVAPPEVDLADATVNVDMPPISD QVVSTPLPDEXDN

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1060	sp_203331z11568	#N/A	-0.854233	#N/A	uncharacterized protein	#N/A
1061	sp_203380_qtyw	#N/A	#N/A	-0.86752	squalene epoxidase 3 (SQE3); FUNCTIONS IN: squalene monooxygenase activity; INVOLVED IN: response to jasmonic acid stimulus, response to wounding, sterol biosynthetic process; LOCATED IN: endomembrane system, integral to membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Squalene epoxidase (InterPro:IPR013698), Fumarate reductase/succinate dehydrogenase flavoprotein, N-terminal (InterPro:IPR003953); BEST Arabidopsis thaliana protein match is: squalene epoxidase 2 (TAIR:AT2G22830.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MFFDQYIFGGVIASLLAVAFLWTLRSTKKREDKY SQEDDVSGVTTSPVGECRSDVTDSDVIIVGAGV AGAALAHTLGKDGRRVHVIERDLTEPDRIVGELL QPGGYLKLIDLGLEDCVENIDAQRVFGYALFKDG KNTKLSYPLEKFHDDVSGRGFHNGRFIQRMREK ASSLPNVQLEQGAVTSLVEENGIIKGVRYKTKAG KELCAYAPLTIVCDGCFSNLRRSLCTPKIDQPSCF VGLVLENCELPYENHGHVVLGDPSPILFYPISSTE VRCLVDVPGEKVPSVANGELAKYLTTVVAPQIPP QLYNAFVAAIDKGNIKSMPNKSMPAAPLPKPGAL LMGDAFNMRHPLTGGGMTVALADIVLLRDLLRPL KDLNDAPNLCRYLESFYTLRKPVASTINTLAGALY KVFRASPDQARKEMRQACFDYLSLGGIFSNGPIS LLSGLNPKPLSLVLHFFAVAIFGVGRLLLPFPSPK RLWIGARLISGASGIIFPIIKAEGVRKMFFPAIVPAY YRVPPVKS
1062	sp_203681z11614	#N/A	-0.870231	#N/A	uncharacterized protein	#N/A
1063	sp_203690_pchy	-0.874157	-0.91469	#N/A	UBP1-associated protein 2A (UBA2A); CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: RNA-binding (RRM/RBD/RNP motifs) family protein (TAIR:AT2G41060.2).	MAKKRKLSAKSAETATSAQQKKLQKKTQNEEEQ EQQPEAQKQPEQEPEKLKAVEAEEEQPPPMEE EVEEEVEEEVEEEEEAMFEEVEEEDEEEGDEEE GDEIEETEEEPKPIEEDHTTTQGSSDDEPIPLLLE PFTKEQLIMILCEAAINHPDVASRIRKVADEDTVH RKIFVHGLGWDTTADTLRSAFSQYGEIEDCKAVT DKISGKSKGYGFILFKSRSGARKALEMPQKRIGG RMTACQLAAVGPVPTSSMGTAAVVQPQQPMSE YTQRKIYVSNVSADIDPHKLSAYFAKFGEIEEGPL GLDKHTGKPKGFCLFVYRSVESAKKALEEPHKH FEGHVLHCQQAIDGPKPGKQHMGHQPRGQHFG GRVDNPNLAGAIGGPPATGAGLMGQSIGGVGYN QGPAAPGLNPALGQALTALLASQGAGLGLNNLL GTLGGGMAGAPGASHGYGAPNVGVNPSVLGGY GQGGLQGAYPNQHMGQGGAGGRSSQHGHMG GVAPYGGH

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1064	sp_203710_mgoq	-0.872068	-0.893589	-0.95182	N2,N2-dimethylguanosine tRNA methyltransferase; FUNCTIONS IN: RNA binding, tRNA (guanine-N2-)-methyltransferase activity; INVOLVED IN: tRNA processing; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; CONTAINS InterPro DOMAIN/s: N2,N2-dimethylguanosine tRNA methyltransferase (InterPro:IPR002905); BEST Arabidopsis thaliana protein match is: N2,N2-dimethylguanosine tRNA methyltransferase (TAIR:AT3G02320.1); Has 1017 Blast hits to 963 proteins in 359 species: Archae - 255; Bacteria - 70; Metazoa - 198; Fungi - 150; Plants - 103; Viruses - 0; Other Eukaryotes - 241 (source: NCBI BLink).	MGSSKDGSVDEEQPKGNNPTSVDFNDYAVIKEG EAEILMHRNNQVFFNKTQVNNRDISIAVLRTFINK RKEEHEAFLSRKAKGAAKTSEEKSDTPASEDVT MESKADSHTSNGIHDTNATDVKPEDASCGTAAE PAQPSNTPGGKFQRELKPPRVLEALSASGLRAL RYAREVEGLGQIVALDNDKASVEACQRNIKFNGS VASSKVESNLADARVYMLTHPKEFEVVDLDPYG SPSVFLDSAVQSVADGGMLMCTATDMAVLCGG NAEVCYSKYNSIPLKSKYCHEMALRILLASIESHA NRYKRYIVPVLSVRMDFYIRVFVRIYSSASAMKET PLKLSYVYQCVGCDSFHLQPLGRAVSKNNNMKY YPGYGPVVPKECADCGKRFNMGGPIWSAPIHDQ DWVTSMLSDVKAMKDRYPAYNRIIAVLTTVSEEL PDVPLFLCLHNLCGTLKCTSPSAVIFRSAVLNAGY RISGTHASPLGLKTDAPMSVIWDIMRCWVKNHP VKPQPADQSGTVILSKEPELQANFTRAVASLSKA QAKKVARFLPNPERYWGPKVRGGRQITSKHVSL MGPEAINGSVNHEEGEESDAKRQKTDHHTSDS

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
						MGFGVRSKRWKRHYLGRLFGFFSFINISEGREE
						VDDVEHSSRNLVEVAVLMNIVQRLHKEWQCPSQ
						KLSVCVVSSYKAQVKEIERRLGNRYEKLNNFELK
						VKSVEEYHEGGEADVVIISTVFRSNSAGSVGSNLD
						RQVANVALNSSRHCLWILGNKKNLVEADSVWED
						IVSDAKERKCLFNAEEDEDLAKVIIDVKSKLDELE
						DLLDKDSNIFKAARWKVVFSENFVKSFAKLPSAF
						SKKLVVNFVLKLANGWRPKSKKINMVCKSSVGIA
						KQFKVEGRYIICTNDIIRVPGYATQILKVWDILHLE
						DVPALVNKLDIIYGALSDDFLICCKESCLQGNLEV
						PVRYASFDYARYKDKNGGSTGSLDAVNGGRSFI
						EKAKVKESLLLMKFYPFSAGVLSHLLSASDGTVL
						NVPFEVSDQEREIILCEKSSFILGRSGTGKTTVLV
						MKLFQKEQQHHIASEGFNVHDALMRGGPSNEPG
						RIGTSLCQLFVTVNPKLCFAVKQHINNLRSIAELFL
						LYYRFRFVHGGNSGEAYISSDVNEMDEYEFKDV
						PDSLEEIPSHVYPLVITFNMFLLMLDRTLGLSYFD
						RFPDIRKIYSLETASSRLVAQAVRRKEVTYDKFRT
						SYWPHFNLQLTRNLDPSRVFTEINSCIKGSLHSR
						ESDDIKLSLASYLQLSKRRVSNFSEDEREMIYKIF
						EDYEKMKVVRGEYDLADLVNDLHHRFKCEKYEG
						EFMDFVYIDEVQDLTMKQLGLFKYICSNVEEGFV
						FAGDTAQTIARGIDFRFEDIRCLFYTEFLHGEERK
					P-loop containing nucleoside triphosphate hydrolases superfamily protein; BEST Arabidopsis thaliana protein match is: P-loop containing nucleoside triphosphate hydrolases superfamily protein (TAIR:AT1G65780.1); Has 6257 Blast hits to 5371 proteins in 1056 species: Archae - 197; Bacteria - 1787; Metazoa - 1336; Fungi - 1200; Plants - 672; Viruses - 106; Other Eukaryotes - 959 (source: NCBI BLink).	KDKGLISATSHLSQNFRTHAGVVNLAQSVIDLIYH
						FFPNSIDPLNPETSLLCGELPILLDCATREDAIIKIF
						QGKGSHGDNSISFGAEQVIMVRDDYARDKLANKI
						GRQALVLTIFDCKGLEFEDVLLYNFFGSSPLKEQ
						WRIIYEYMNEKLCLLDPETFPSFTNAKHDVLCYEL
						KQLYYAITRTRQRLWICEDRDGFLSPMADYWQK
						LNLVQVEILDDSFVQAMQVASSLEDWKWRGMK
						MLEVRNYKAATQCFERARDFYWEKFARASDLKE
1065	sp_204240_dpim	-0.852116	#N/A	#N/A		TANNWRGLRLDKSLEMLREAAEIFESIKNIKKAIE
						LGQIDAPAAQHFNDLYKAVQSSMDLYYVYRWNQ
						MQLTQNNIDIAELKKRAEAAESALKINEKKLESAS
						SNLEAANRRLEEVEPKLKAETERADGLTEELTDL
1066	sp_204260_ksmi	-0.974276	-0.930711	-0.998103	uncharacterized protein	NQSLPVVKKTAAKRAVDKLLQSDWFNDILTLRHN
						GGWCAAHRVVCHXSSGRLPCEEAG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1067	sp_204310_synh	0.5957285	#N/A	#N/A	Peptidase M1 family protein; FUNCTIONS IN: metallopeptidase activity, zinc ion binding; INVOLVED IN: proteolysis; LOCATED IN: apoplast, chloroplast, plasma membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Peptidase M1, membrane alanine aminopeptidase, N-terminal (InterPro:IPR014782), Peptidase M1, membrane alanine aminopeptidase (InterPro:IPR001930), Peptidase M1, alanyl aminopeptidase (InterPro:IPR012779); BEST Arabidopsis thaliana protein match is: aminopeptidase M1 (TAIR:AT4G33090.1); Has 7901 Blast hits to 7867 proteins in 1782 species: Archae - 112; Bacteria - 3789; Metazoa - 1713; Fungi - 432; Plants - 227; Viruses - 0; Other Eukaryotes - 1628 (source: NCBI BLink).	MARLVLPCKGPALSKTNFLGLVSSSPLQALSRVN CVQRPVKNLTRFRKFLSLETSLRRNCHFSYPALY RSQEVNRRLICSVATEPSPKQAEDLNMEAPKEIF LKDYKKPAYFFDTVDLKFLLDEDKTHVFSKISVFP ADEASASSPLILDGSDLKLVSVKIDGKDLNEGEFH VDSRHLTLTSPPTEKFTLEIVTEIYPQNNTSLEGL YKSSGNFCTQCEAEGFRKITFFQDRPDIMAKYTV RLEADKTKYPVLLSNGNLVGQGDIEGDKHFAIWE DPYKKPCYLFALVAGPLESRDDTFVTRSGRNVTL RIWTPADDLSKTAHAMYSLKASMKWDEDVFGLE YDLDLFNIVAVPDFNMGAMENKSLNIFNSKLVLAT PETASDADYAAILGVIGHEYFHNWTGNRVTCRD WFQLSLKEGLTVFRDQEFSSDMGSRPVKRIADV SKLRSYQYPQDSGPMAHPVRPHSYIKMDNFYTE VVRMYKTLLGSQGFRKGMDLYFKRHDGQAVTC EDFYAAMRDANSADFANFLLWYSQAGTPIVKVT SSYDAEKRTFALKFSQEVPPTPGQSTKEPMFIPV AIGLLDSHGKDMPLSSVYHDGKLESVAIDGKPVH TTVLRVTKKEEEFVFSDVAEQPVPSLLRGFSAPI RLESDLSDSDLSFLLAHDSDEFNRWEAGQILARK LMLSLVADFQQNKPLHLNANFVDGLRSILCDNSL DKEFIAKAITLPGEGEIMDMMAVADPDAVHAVRT FIKKELALQLKTDLLRTVLENKSSEQYEFNHSNM ARRALKNISLAYLTSLDDPELTELAVNEYKAATNM TDQFAALAAITQNPGKTHDEVLADFYNKWQHEF LVVNKWFALQAVSDVPGNVENIKKLLDHPAFDLR NPNKVYSLIGGFCGSPVNFHAKDGSGYQFLGEL VVQLDKINPQVASRMVSAFSRWKRYDETRQNLA KAQLEKIMSANGLSENVFEIASKSLAA

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1068	sp_204680_agai	-0.899067	-0.901366	-0.898123	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: U2 snRNP auxilliary factor, large subunit, splicing factor (TAIR:AT4G36690.1); Has 62164 Blast hits to 35633 proteins in 1636 species: Archae - 185; Bacteria - 4712; Metazoa - 29770; Fungi - 5811; Plants - 3284; Viruses - 214; Other Eukaryotes - 18188 (source: NCBI BLink).	MSSSTRRKLKSAIICERTLDDTSEESSARTRPLSY DEIMLKRKSKKEVEGSVEEAQDAETSRHDETVR NPEHRTRDTNGASMKNSRDSQRKASPKRVEKG YRKEHTHITSDSRDYQREASPKKVDKSSRMEHT HIKRDSRGSQRESSPRKVEKSYRTERTHIKSDSR DYQREASPRKVEKSSRKEHTHVKSDPRDYQREA CPRKKADKSSRTEHTHVKSDPQDSQREGSPRK KVEKSSKTEHADIKSDPRNSQLEASPRKVEKSSR TERTHSKSENKERRDVECRSKPKHDNDASNATE SKTDRQGGMRSKIEERSRSDARYTNENKHHRDS VAQDRYVERNGGASERENKRRDRSGDDEKHRM GDAVKRHDSGSRRNSEISERKENMELSHSRHDE TRSKRRRSRSQERTKDKARRSTSSSPKAQKTVL VGHVLDHGDTSTHSERDRSRRHSDVDRSKMSS DGSSSHYRRQGSSSGLGGYSPRKRKSEAAVKT PSPIHSPERKSAGWDLPHFAADKNLSTSVPSGT QLVSQTKPSNMIELAAGFSVASSMGKPTSAIFSN SLSAFKQATIDSVQLTQATRPMRRLYIDNIPGSAT EKDIMECFNSFLLSSGSNYIRGTRPCISCMIHKEK GQALVEFLTPEDASAALSFDGRSLSGSILKIRRPK DFVEVPSGSSERLKAGAVSNRGDDSVREVSITAI SDFVEDSSHKIFVGGIAEGLSSEMIVEIVSAFGTL KAYHFEVNEDLKEPCAFLEYADQSVTLNACASLN GMRLGGRVLTVAQAVPNTQSVESSRISPFYGIPE HVKPLLEMPTQVLKLSNVLDQKTMDSISETDLEEI VEDVRLECSRFGSVVSINVVKGGKTHASISEAKE VTYDVNTIQPPEQIIHQLDMNPVEVAADHYDRDG SNVEPPTDVKEFDAIGGVESDKHELVPATTDFEI GAAGTSIFEESEEADKISPQNADDEAAVIDDSAID DMARVNDSEENRQEVSIQDSGVSSDLNPNGSM GIEKSGDGDTVHAEDVPELDKNNDTIQEVVPLPT NLGSSEEGIKDEDSYDRGNIFEVGCVLVEFRRPE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1069	sp_205050_ixnw	#N/A	-0.9111345	-0.9863945	serine hydroxymethyltransferase 6 (SHM6); FUNCTIONS IN: pyridoxal phosphate binding, glycine hydroxymethyltransferase activity, catalytic activity; INVOLVED IN: glycine metabolic process, L-serine metabolic process; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Pyridoxal phosphate-dependent transferase, major domain (InterPro:IPR015424), Serine hydroxymethyltransferase, pyridoxal phosphate binding site (InterPro:IPR019798), Pyridoxal phosphate-dependent transferase, major region, subdomain 1 (InterPro:IPR015421), Serine hydroxymethyltransferase (InterPro:IPR001085); BEST Arabidopsis thaliana protein match is: serine hydroxymethyltransferase 7 (TAIR:AT1G36370.1); Has 11545 Blast hits to 11518 proteins in 2846 species: Archae - 258; Bacteria - 6402; Metazoa - 340; Fungi - 287; Plants - 348; Viruses - 6; Other Eukaryotes - 3904 (source: NCBI BLink).	MDLSHPQSNLSLGFLSHVSAPASSIPANSNPNRI LINDDSISFQIDSRLRESSIPMPSIPLQLLESQAVQ NCRKIENGCVGERERVAVVEPKRDVGEDAEERE EENRGEDEDEEVEEFRILGHSMCLKRRRDREDS VSSSTALAKRVFFDSSMEARKGAVRAWGNQPL RLADPEIFNIMDKEKLRQCKGIELIASENFVCRAV MEALGSSLTNKYSEGMPGQRYYGGNQYIDEIELL CCDRALAAFGLSSESWGVNVQPYSCTSANFAVY TGLLLPGDRIMGLDTLSGGNSSHGHYTPSGRKV SGASIFFESLPYKVNPQTGYVDYDKLEEKALDFR PKILICGGSSYPREWDYARLRQIADKCGAILMCD MAQISGLIAAKECVSPFDYCDIVTSTTHKSLRGPR GGIIFFRKGKKPRKRSSLNPSEDNEQYDFEDKINF AVFPILQGGPHNNHIAALAIALKQVASPEYKAYMQ QVKRNAQALASALLRRKCKLLTGGTDNHLLLWD LRPLGLTGKNFEKVCESCHITLNKVALIGDNGAIT PGGVRIGTPAMTTRGCLESDFEVIADFLIRAAQIT SIMQREHGKVAKCFIKGLQSNKDIIDLRNRIEAFA SQFAMPGFDV

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1070	sp_205490_oaoz	-0.95359	-0.914764	-0.944184	EMS-mutagenized bri1 suppressor 1 (EBS1); CONTAINS InterPro DOMAIN/s: Glycosyl transferase, family 8 (InterPro:IPR002495), UDP-glucose:Glycoprotein Glucosyltransferase (InterPro:IPR009448).	NVQTALRAKWEGTPLLLEAGELLSREWKDLFWD FVDVWLQHENDDVDSNTAKGCMLKIVKHGRSFL SEHLGSMFEFSLTLRSASTRLVLYRQLAQESLSS FPLTDEAVMNAGDVASDLHGNTNIGNRKPDPLLV GVNLKSPGGKCCWVDVGASLFFDVAELVEWLQI CGGDSSQQPELFEFDHVYSVSNGSSPVAILYGA LGTDCFREFHGVLVHAANKGKVKYAVRSVLPIGC EVKTSRCGALGTKSPVNLGGYGVELALKNMEYK AMDDSMIKKSITLEDPHTEDLSQEVRGFIFSKILE RKPGLLSSEIMAFRDYLLSSSISDTLEVWELKDLG HQTAQKIVHASNPLQAMQEINQNFPSVVASLSR MKLNDSVKDEILANQRMVPPGKSLMALNGALINI EDIDLYLLIDLVHQELSLADQFSKLKIPEPTIRKLLS TVPPSESSMFRIDFRSSHVHYINNLEEDSLYKRW RSNVNEILMPTFPGQLRYIRKNLFHAVYVIDPATS CGLEVIDMANLYENQFPMRFGVILYSTTFIEKVEK NGGENLADFSVSSNDEDISSLIRLFIYIKEHHGML TARKFLGNVHSEFGDSGEEAPERHHVEGAFIDAV LSKSKSPPQDVLLKLDKEKDFKDVAQESSMSVFK LGLSKRKCCLLLNGLVYNSNEETLLNAMNDELPR IQEQVYYGNIDSHTDVLDKFFSENGIRRYNPLIITD GKFKPKFVSLSNSILNERSALHEINYLHSPETIDDL KPVTHLLAIDLTSTKGMKLLQEGIRYLEDGSTKAR IGVLFVPNFAFDSPGLLFVKIFQVTSLSYSHKEGV LKFLDQLCSLYEHVFMKHSSQQAESMKAFIEKVS DLADVNGLPSKGFSSLLTEFAAEDLQNQLNKVAL FLHKHLGHESNSNAVITNGRVTIVSDENTFLSHDL NLLESFEFKQRINHISEIIEDLKWENTDPDSLTSKF ISDVIMLVSSSMATRDRSTEGARFEILQATYSAVV LSSESPIHIDAVIDPLSPSGQKLTSILQILGKYFQP NMRIVLNPLSSLVDLPLKNYYRYVVPSMDDFSCS
1071	sp_205590_jons	#N/A	-0.931214	-0.957779	Pre-rRNA-processing protein TSR2, conserved region; CONTAINS InterPro DOMAIN/s: Pre-rRNA-processing protein TSR2, conserved region (InterPro:IPR019398); BEST Arabidopsis thaliana protein match is: Pre-rRNA-processing protein TSR2, conserved region (TAIR:AT3G22510.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MDEPKKPQVLPTVAIPHFQEGISLVFSKWWALQ MAVQNEWGGHDSHTKSLNFSSDIFNFFYKPNRK EPIYIDDLEEELNVGLDSFNTTADDGSVEEVAVQL MVMYEECLEGNYQSIQKLKQMKSQPVLRVEQDA SDSDEGEDENENSMGNDEASNMMVDTPESFLS TNSTETQKNNSRHQQMEEAEDGWTVVPSRKSK G

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1072	sp_206670_eafq	#N/A	-0.875383	#N/A	proteophosphoglycan-related; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 7 growth stages; Has 4120 Blast hits to 443 proteins in 120 species: Archae - 2; Bacteria - 57; Metazoa - 120; Fungi - 93; Plants - 84; Viruses - 11; Other Eukaryotes - 3753 (source: NCBI BLink).	MARTNKYTSLNFNDIYDKKILTATATATSRSSSSS AAHPHKSYSLSASSRLGHGGMLVLSRPSPKPQP PPQQPPPQSQPQQPPSNSVAPSATPSPRSSKS PDSDSISLRPQGKTGGSPTFQERENGNNKVDVP SPKSNKFVPPHLRPGFVAREAASVPEFQKPRQG NFRPGSAHGNYHGSPNRYIEDGRPKSGSGYERR MRAAGPVHGAAHAHAPAPVGVESDSIVSNRPPS GGSRPCSSGXCIYKSWTLFPIVELEGSIVADIFQG
1073	sp_206711z12081	#N/A	-0.858462	#N/A	uncharacterized protein	#N/A
1074	sp_207030_ucfs	#N/A	-0.861386	#N/A	Integrase-type DNA-binding superfamily protein; FUNCTIONS IN: DNA binding, sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent; LOCATED IN: nucleus, chloroplast; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: DNA-binding, integrase-type (InterPro:IPR016177), Pathogenesis-related transcriptional factor/ERF, DNA-binding (InterPro:IPR001471); BEST Arabidopsis thaliana protein match is: Integrase-type DNA-binding superfamily protein (TAIR:AT2G40340.1); Has 5061 Blast hits to 5054 proteins in 220 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 5051; Viruses - 0; Other Eukaryotes - 10 (source: NCBI BLink).	MNARRRRNGYNSVAEILAKWKIQNQQYGSTIEE RKKNTQKGPARGSRKGCMPGKGGPENLGCTFR GVRQRTWGKWVAEIREPISSSATGYGNTTQSKR SRRLWLGTFSTAIEAALAYDEAAKVMYGPSAILN FPSQQNVEHEDQALCGSNMQSTTASESTEGSE NYKGFNEELKSPCASQVVDVGGSKEEFGMGSV NSNGGSYLNLGFDYTHDWVMEPLDGFGGVNFG LKQEDDYEFDQNGLFYSDSSDSMANYKLQQPLS GPLGIDFITSSDFVRSNTNLDGLH
1075	sp_207310_ykpd	#N/A	-0.857212	#N/A	uncharacterized protein	MATLFSYILQDTVSYAVDSHSKYFPQLEHLSWKM QLYHGFLQLDAHCVEGLITRMP
1076	sp_207340_yhrz	-0.861371	-0.848215	#N/A	HSP20-like chaperones superfamily protein; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20 (InterPro:IPR002068), HSP20-like chaperone (InterPro:IPR008978); BEST Arabidopsis thaliana protein match is: HSP20-like chaperones superfamily protein (TAIR:AT1G59860.1); Has 7238 Blast hits to 7237 proteins in 1623 species: Archae - 225; Bacteria - 4290; Metazoa - 129; Fungi - 318; Plants - 1626; Viruses - 0; Other Eukaryotes - 650 (source: NCBI BLink).	MSLIPRNTSNNQTPFTPSQNFWDHFFSHDFPL QPFTNLPLSPFLTPPHSARTVPYSELAVETFGEV RTRLECKEIPEAHLIVADLPGIEKHEVEVAAEDGG FIKISAVDGRFSWRLPEGAEVHLMSSSVENGV LTVVVPKFVGEMMWERENNEGGRDVRVIEITGE DE

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1077	sp_207930_dpjf	-0.967765	-0.965412	-0.918818	Pentatricopeptide repeat (PPR) superfamily protein; CONTAINS InterPro DOMAIN/s: Pentatricopeptide repeat (InterPro:IPR002885); BEST Arabidopsis thaliana protein match is: Pentatricopeptide repeat (PPR) superfamily protein (TAIR:AT3G62470.1); Has 32572 Blast hits to 10552 proteins in 258 species: Archae - 4; Bacteria - 9; Metazoa - 159; Fungi - 256; Plants - 31316; Viruses - 0; Other Eukaryotes - 828 (source: NCBI BLink).	MWRTCVILIVRQCKNSIYFSNLQNPIFYSVSRAYE TSIFGFGVIPKLHSSKGCFEFGNPLRKFYSTDDFV GEDENNLVYSNVDGVYRIIMENVGSDKNMEKAL EQARIPLTTELVIETLDRLRFEEKIAFRFFMWAAR QEHYHHEYKVYNEMIDILSTTKFKIKQFKIVCDML DYMKRNDKTSIPIEVLLTILRKYTEKHLTKVLTCSK KKRGVRVKMQPEIKALNLLLDALCKCNLVEDAEIL FKKVKNKVAPDANTYSILFFGWCRVKNPARGMKI LEEMSSMGFTPDSFTYNAAIETYCKAGMVTEAV GLFELMKTKPSISSPPTAKTYAIVILALVKIDRIDDC FRVIGDMIQSGCLPDVSTYKELIQGMCLAGKVKE TYWFLEEMEKKGYPSDIVTYNCFLKVLCENRKAD EAIQLYRRMVSVDCVPSVHSFNMLIEMFFSMGDP VGALETWDEMGKKSCERDVNTYCVMIEGLFVCD QIENACLFLDDVVHQGIKIPYLKFDLFLRHLSRIDD LRRIHKLSEHMKRFFNPAMARHFVLNQKRKELSY RNK
1078	sp_208590_fpkc	#N/A	#N/A	-0.894713	nucleolar essential protein-related; CONTAINS InterPro DOMAIN/s: Ribosomal biogenesis, methyltransferase, EMG1/NEP1 (InterPro:IPR005304); Has 1079 Blast hits to 938 proteins in 280 species: Archae - 143; Bacteria - 12; Metazoa - 353; Fungi - 181; Plants - 69; Viruses - 2; Other Eukaryotes - 319 (source: NCBI BLink).	MGKTKRKAGSLQPCSTNRKKLNRDGLAVKGDEE ELGEKELNSIQKFSHPTVNEVKNEDDEVNLCGIP LSLTCKDAGAGAGVTFVLEKASLTLANIGKRYQIL NSNEHASYLRKKKMDPYKYRPDIVHEALNDIMNS RLSKSGRLKAVYIRTEDGVLIKVEPYAKIPSDLES FCSMMAELLQTLSIKAKGNRKKLLRVVANPVTKY LPPSSPKIGLSFSSEKAVELREHVSGINRDEDVVF VVGVMAHGKIECDYIEDLVSVSPFHLSAAMCLRII LIAMERKWSIL
1079	sp_208960_hekk	#N/A	-0.850168	#N/A	uncharacterized protein	MHCRCLINAFDSKAPAKYSEETADTAGKLIIQGTV ILQKKKTMSDRLFSSILKIYVQLVTDTGLQLIN
1080	sp_209060_ejtg	-0.980783	-0.968461	-0.990816	uncharacterized protein	MVLSVQSMPGGLENIACDGKPSAIPRDVSPNEKI QEVLHEINVNEIANETAVQCEEPRIVDAITDPTQQ TNAGTTLEVDVQPNLSEQPV
1081	sp_209250_hzzy	#N/A	-0.877626	#N/A	ARF-GAP domain 7 (AGD7); FUNCTIONS IN: ARF GTPase activator activity, DNA binding, zinc ion binding; INVOLVED IN: regulation of ARF GTPase activity; LOCATED IN: cytosol, nucleus; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Arf GTPase activating protein (InterPro:IPR001164); BEST Arabidopsis thaliana protein match is: ARF-GAP domain 6 (TAIR:AT3G53710.2); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	MAGASRRLRDLQSQSGNKTCVDCSQKNPQWAS VSYGIFMCLECSGKHRGLGVHISFVRSVTMDSW SEIQLKKMEAGGNDKLNMFLAQYGIPKETDIIPKY NSNAATVYREKVSALAEGRSWRDPPVVKESVG GGSRKPPLGGGGGSGGRSSGNGNSGGWDSW DNDDSFRSSSNGGGNDMRRNQSTGNFRNNNGI GNGNGNHGPHKSRSSSDFTRSQLEESAANKDS FFARKLAENETRPDGVPPSQGGKYVGFGSNPNP NPTPR

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1082	sp_209690_axnw	-0.853681	-0.900122	#N/A	LOB domain-containing protein 37 (LBD37); CONTAINS InterPro DOMAIN/s: Lateral organ boundaries, LOB (InterPro:IPR004883); BEST Arabidopsis thaliana protein match is: LOB domain-containing protein 38 (TAIR:AT3G49940.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	MSCNGCRVLRKGCSESCILRPCLQWIDTPESQG HATVFVAKFFGRAGLMSFISNVPDSQRPALFQSL LYEACGRTVNPVSGAVGLLWTGNWHLCQAGVE TILRGGTLRPISEFLVGAPPSPIPTSVETDEISEVF CPPKPNTTTAATNPNPNTWRKQIGVNMMKQDLN GNYNNNKRSRSSPPPKFKKEEEDYKQLDLSLT PPTFLHAPSQQRPGTPSLNSEESAATTVSFESD NYNYQPKLLNLFH
1083	sp_209880_xonn	-0.926102	-0.876388	-0.923632	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein; FUNCTIONS IN: peptidyl-prolyl cis-trans isomerase activity; INVOLVED IN: protein folding; LOCATED IN: plasma membrane; EXPRESSED IN: 26 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Cyclophilin-like (InterPro:IPR015891), Peptidyl-prolyl cis-trans isomerase, cyclophilin-type (InterPro:IPR002130), Peptidyl-prolyl cis-trans isomerase, cyclophilin-type, conserved site (InterPro:IPR020892); BEST Arabidopsis thaliana protein match is: rotamase CYP 1 (TAIR:AT4G38740.1); Has 16606 Blast hits to 16571 proteins in 2692 species: Archae - 108; Bacteria - 6886; Metazoa - 2917; Fungi - 1384; Plants - 1286; Viruses - 4; Other Eukaryotes - 4021 (source: NCBI BLink).	MFDLSYGRKHVVFGKVVEEMKVVKKMEQFGTSE GRPTEQVKIVDCGEVSSGKNESAVMPEKGKVKK LKKLPSSDDNSDGKGRGRRKRSPKDRRRKKRK RSQRNSSSSSDQSDSESDSSGSESNASLSDFSS SSDGRRVRKRSKRSRRNRKKSGHNSERSPESL RASDSETTSSKSGSEEEKSRHRTSSKLKKKSSIP TEEQQQKAHVQKVVINPPEVEDEFSRLTERIEKL RPCNWSPIVFFNIWYHGKFILHPKKGYEQGLSEM LLIDRERLSFIHLKKVIHMVGCRKKYTIWFVADIEK LHEEGLAIQDIADFYDLVELTKSNTVVHIVQHFDIE EESKGAVKRTVQPPPSTKKPRRSTGSCKVKGGNT SMHSWGSVEPAPCVRKIPGMGCAVRQFATLLSN LTVKQKMWVVDMGFEELLFMRIGNIDARFGYGIT CRFDPNTLELEIRDNYKVKIDEELVGLVLGLRSGH HLLFPSYSNKSTALNNKKLYKQNTMYGAITESEV YGKIVSEPQDKETFQKHFLLYALGTILRPTQKRG WISPLHASVLYLASESADYNWARYVLDWLVKYG KKFNSSKEGYGGCTLLLLIYIDRLTSTRVNLNIIPN SCRLRVLDQESVDRVLKSDQGPNGDYGYGEVV EDYYYGDM
1084	sp_209910_xecp	-0.85081	-0.940359	-0.870088	RNA-binding (RRM/RBD/RNP motifs) family protein; FUNCTIONS IN: RNA binding, nucleotide binding, nucleic acid binding; INVOLVED IN: biological_process unknown; LOCATED IN: nucleus; CONTAINS InterPro DOMAIN/s: RNA recognition motif, RNP-1 (InterPro:IPR000504), Nucleotide-binding, alpha-beta plait (InterPro:IPR012677); BEST Arabidopsis thaliana protein match is: U1 small nuclear ribonucleoprotein-70K (TAIR:AT3G50670.1); Has 272 Blast hits to 272 proteins in 16 species: Archae - 0; Bacteria - 0; Metazoa - 2; Fungi - 0; Plants - 270; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MSGTNRSSNSVNAVFYAQAYHPIQAGSIDGTDIL PHDNAVYRALLCSSAGLYDPFGDPKVVGDPYCT LFVGHLSRSTTEDTLCKAMSRYGKVKNLRIVRDI VTGASRGYAFIEYEREKEMRHAYEDAHHSFIDDS EIIVDYNRQQLMPGWIPRRLGGGLGGKKESGQL RFGGRERPFFRAPLRPIPHDDLKRLGIPPPPEGRF MTRFQVPSPPRRKRNSEERETSPFRRKRNYYEE KETSPPQRKRYSEDQETSSSRRERNSKDWETSL SLSSSSRRERNSEDREESSPRRKRSSQDRESRR DKHSHKHQKHRRHSNHHEKRSRSREY

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1085	sp_210200_jgiq	-0.893409	-0.869536	#N/A	MAGO NASHI (MAGO); FUNCTIONS IN: protein binding; INVOLVED IN: pollen tube guidance, sex determination, embryo development ending in seed dormancy; LOCATED IN: in 6 components; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Mago nashi protein (InterPro:IPR004023); Has 476 Blast hits to 476 proteins in 200 species: Archae - 0; Bacteria - 0; Metazoa - 223; Fungi - 96; Plants - 76; Viruses - 0; Other Eukaryotes - 81 (source: NCBI BLink).	MAGEEDAGEFYLRYYVGHKGKFGHEFLEFEFRP DGKLRYANNSNYKNDTMIRKEVFLTPAVLRECRR IISDSEIMKEDDNLWPEPDVIGRQELEIVMGNEHI SFTTSKIGSLVDVQSSKDPEGLRIFYYLVQDLKCF VFSLISLHFKIKPI
1086	sp_210310_jmun	#N/A	-0.85122	#N/A	SAUR-like auxin-responsive protein family ; CONTAINS InterPro DOMAIN/s: Auxin responsive SAUR protein (InterPro:IPR003676); BEST Arabidopsis thaliana protein match is: SAUR-like auxin-responsive protein family (TAIR:AT1G72430.1); Has 128 Blast hits to 128 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 128; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	MAKAGKLTKLRSALKKMNSFSKLRRENSINCTPA PSDDFHVSSTNNHNNVGGLLHTVYVGKSLRRYQ VTSEVLDNPVIQELVDRSNGCDEETTVIGCEVVL FEHLLWMLQNIDSQSESLEELVGLYAY
1087	sp_210810_kdsc	#N/A	#N/A	-0.849965	HAT22; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity; INVOLVED IN: regulation of transcription, DNA-dependent, response to cytokinin stimulus; LOCATED IN: nucleus; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Homeobox, conserved site (InterPro:IPR017970), Homeobox (InterPro:IPR001356), Homeodomain-like (InterPro:IPR009057), Leucine zipper, homeobox-associated (InterPro:IPR003106), Homeodomain-related (InterPro:IPR012287); BEST Arabidopsis thaliana protein match is: Homeobox-leucine zipper protein family (TAIR:AT2G22800.1); Has 6652 Blast hits to 6630 proteins in 455 species: Archae - 0; Bacteria - 0; Metazoa - 4237; Fungi - 277; Plants - 2018; Viruses - 3; Other Eukaryotes - 117 (source: NCBI BLink).	MDSHSHSKYQIQDQALTLGLGISSSSSLLSSKTK QVQVSIPSAATSSCYHHHEPSLTLSLSGDTYTTD TPIATPVKVVEPPCSGGGGVSVVSVKRERESAC SSDDLEIERGGGSSRISDEDEDGSAGPTNTPRKK LRLTLDQSALLEESFKQQSTLNPKQKLALADRLN LRPRQVEVWFQNRRARTKLKQTEVDCEHLRKCF ETLTDENRRLQKELQELKALKLGQPFYMHLPTAA AAALTMCPSCERIGASVPTIQGTPTTTSTSTTTNA SNKGTPFTLGLSKSHFYPSPHQIPSFTC

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1088	sp_210890_yuic	#N/A	-0.8580856	-0.8550231	high mobility group (HMG); CONTAINS InterPro DOMAIN/s: Structure-specific recognition protein (InterPro:IPR000969), High mobility group, superfamily (InterPro:IPR009071), High mobility group, HMG1/HMG2 (InterPro:IPR000910); BEST Arabidopsis thaliana protein match is: high mobility group B1 (TAIR:AT3G51880.2); Has 8066 Blast hits to 7188 proteins in 653 species: Archae - 0; Bacteria - 14; Metazoa - 5623; Fungi - 822; Plants - 671; Viruses - 15; Other Eukaryotes - 921 (source: NCBI BLink).	MADHLFNSISLGGRGGTDVATLTTFFRSSFGITP EERQVSVSGHNWGEVDFNGNMLTFLDGSKQAF ELSLTDVSQTQMQGKTDAILEFHLDDAAVANEKD TLMEISFHIPNSNTQFVGDENRPSAQVFHDKIMS MADVDAEEAVVTFDGIAILTPRGRYSVELHLSFLR LQGQANDFKIQYSSVVRLFCLPKSNQPHTFVVITL DPPIRKGQTLYPHIVMQFETDYVVESTLSMNEDL LNTKYKDKLESSYKGLIHDVFTNILRALSSAKVTR PGKFRSCQDGYAVKASLKAEDGILYPLEKSFFFL PKPPTLILHEEIDFMEFQRHVVAGSNMHYFDLLV KLKTDQEHLFRNIQRSEYQNLFDFISAKGLKIVNV GEGANLPSISIAPLSSDDEVVDPHLLRRIRDAQDES DEEDEDFVIDKDDGGSPTDDSEEDHSDGSVSGD EKEKPAKKEKKDSASLKPSSSKKKSRDGDEDGS KKRKTKRKKDPNAPKKAKTGFIFFSQAERENTKK ENPGIAFLEMGKALGEKWRIKMSAEEKEPFEAQA RADKKRYSEEIGDYKNPQSVNVDSGNDSESE
1089	sp_210970_tnsk	-0.8603036	#N/A	-0.8729934	uncharacterized protein	MNRVSSHRAGEPSRRRAPPKKRRSHRFHHLNSS RAPSKPDSAIEHRRLGEPSRRRRAPPKKRRSRAPRI EKARFSKTATVTRE
1090	sp_211620_ooug	-0.9113359	-0.8693318	#N/A	HXXXD-type acyl-transferase family protein; FUNCTIONS IN: transferase activity, transferring acyl groups other than amino-acyl groups, transferase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: synergid, pollen tube; CONTAINS InterPro DOMAIN/s: Transferase (InterPro:IPR003480); BEST Arabidopsis thaliana protein match is: HXXXD-type acyl-transferase family protein (TAIR:AT2G40230.1); Has 2626 Blast hits to 2616 proteins in 147 species: Archae - 0; Bacteria - 2; Metazoa - 0; Fungi - 66; Plants - 2554; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	MTADTKKSPEIPDCYYPNSPLLITPDLPTPKHSLY LSNLDDQNFLRFSIKYLYLYEKGICIENLKQSLSKI LTEYYPLAGRLRATDESDDQCQKLEVDCNGEGA LFAEAFIDLSVSDFLEFSGKPNNSWRKLLYKVDA LSFLDVPPLVVQ
1091	sp_211890_yiyx	-0.8527396	#N/A	-0.8944867	uncharacterized protein	MPHIKANHLSSLFGSSRRIAPSSSVAAVVKDAVIT SSKKKLRSPHTSSSSYSKVKSSFRKQTRQKNHP PSVADSATEGLFDW
1092	sp_211940_timy	#N/A	-0.8658613	-0.864442	uncharacterized protein	MMAISLIATGHRKIHRGHSDCHLAAAAYRDSRIAA AS
1093	sp_212320_gcmw	-0.8868855	#N/A	#N/A	uncharacterized protein	MHRVAQHVAPEASKDPYVVSLNQDGKTIATPGR MWNGQKPVRRSVRGADRRRLGGRKNPGSPGA SSSRLW

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1094	sp_213330_zpuu	#N/A	-0.865212	#N/A	Nucleic acid-binding, OB-fold-like protein; CONTAINS InterPro DOMAIN/s: Nucleic acid-binding, OB-fold (InterPro:IPR012340), Nucleic acid-binding, OB-fold-like (InterPro:IPR016027), Protein of unknown function DUF223 (InterPro:IPR003871), Replication factor A, C-terminal (InterPro:IPR013955); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G60920.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MKKSNIIELEDLEQKSLSCTLWDEYADEIHTYISK PKDTPTLIIQFGKIKTWEGAMILSSSLFATKLLFN ANIDEVKAFNESVCEGGRATSERLTQTFSQAPYT IEDEFLQSTDRKRIDEIRETQEVRNCVALTTIICIES SYGWXYKVQIRVVDDTGNASFVLFDCEVAQIVG KSAEFLREHYAKDDEGDSIPAELKKMLDKRYLFKI KISEYNLHQNWPVFTVTKMTDDEKLINAFLKSSA SSQEQEEEEETSKANISGGYEDDASKEHENDNT SDICSNSCITPNQKRVFQDLGETNGQCLTSAEMS TSKSRRIIVVKKEKDEI
1095	sp_213730_onwy	#N/A	-0.881478	#N/A	uncharacterized protein	NGTREFQIGIGVGDGVIIEGDKVGINITMAWSRPR FVA
1096	sp_213940_xghg	-0.931413	-0.9728	-0.911363	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages.	FIAFVCTLIYEMIGDLWSSVCSIVHIASTTEATVTT TYEVSMWRSLWNDLFSKVFRAVRSILHGFVTFFT ACNRHRL
1097	sp_214030_ghxg	#N/A	-0.859318	#N/A	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 14 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G31570.1); Has 181008 Blast hits to 85359 proteins in 3551 species: Archae - 3290; Bacteria - 48304; Metazoa - 70793; Fungi - 13943; Plants - 10118; Viruses - 785; Other Eukaryotes - 33775 (source: NCBI BLink).	VCKQLENSRSELSSQVEELQDSLNELKIVADKSV EERDQCNMKVCDLETELAELQNICSELRLKLENY HEIEANLREKDVDISSLQSSLTMKEKEVEEALFW ASHVENLLSRSNKIDVSVEDSQGGDLEGQGSSQ FKKLLYVVDSFPELQ
1098	sp_214230_hjjh	-0.956117	-0.876134	-0.926383	uncharacterized protein	MDEEQSQNQPSECIDLDQDEEFVDLVDDDDLTQ TPSIPSTVQKTSKKPTSEAWEFFEK
1099	sp_214380_wxqe	-0.880249	-0.973415	-0.945812	uncharacterized protein	NGNGNGNGIYFGERCYDEECGVDLVRYSAQRDF SCVRFDACKVVLVPKLCSISKVEVEMLKSYLDSN RSVIISQSGIWEAALRTLAYQLVGDLVSVVAVKT DFRPCLTDIPLTTSALVSEDGLKSGRTGFVFSVIL SQTSLILNNK

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1100	sp_214510_yfkg	-0.822748	-0.984136	-0.940984	detoxifying efflux carrier 35 (DTX35); FUNCTIONS IN: antiporter activity, drug transmembrane transporter activity; INVOLVED IN: anther dehiscence, flavonoid metabolic process, pollen development; LOCATED IN: plasma membrane, membrane; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Multi antimicrobial extrusion protein MatE (InterPro:IPR002528); BEST Arabidopsis thaliana protein match is: MATE efflux family protein (TAIR:AT4G00350.1); Has 11724 Blast hits to 11622 proteins in 2105 species: Archae - 230; Bacteria - 8860; Metazoa - 143; Fungi - 322; Plants - 1354; Viruses - 0; Other Eukaryotes - 815 (source: NCBI BLink).	MNLNGLQAMLLYMVFIGINAAVSIRAQMSSDSSIQGRPDIV
1101	sp_214920_sxec	-0.878919	-0.936753	-0.874734	uncharacterized protein	MNFAAASACVLVETRPGK
1102	sp_214950_mpnd	-0.882363	-0.891706	-0.864942	ISU1; CONTAINS InterPro DOMAIN/s: NIF system FeS cluster assembly, NifU, N-terminal (InterPro:IPR002871), ISC system FeS cluster assembly, IscU scaffold (InterPro:IPR011339); BEST Arabidopsis thaliana protein match is: ISCU-like 3 (TAIR:AT4G04080.1); Has 5453 Blast hits to 5453 proteins in 1998 species: Archae - 110; Bacteria - 3654; Metazoa - 148; Fungi - 154; Plants - 89; Viruses - 0; Other Eukaryotes - 1298 (source: NCBI BLink).	MLAEDAIKAAVKDYEAKRAKQSGSSSEAASTEKAV
1103	sp_215530_teit	#N/A	#N/A	-0.8883849	FUNCTIONS IN: signal transducer activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Uncharacterised conserved protein UCP013022 (InterPro:IPR016607), Regulator of G protein signalling (InterPro:IPR000342); BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF3550/UPF0682) (TAIR:AT3G03570.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MALRNGGVEITQEQQYTTTQDHPSRTFRALVEKADRKFSRVRDLPSFGPDPNPARYLNKVFKAYMRVWKYQRQNRAKLVGSGLQRWEIGEIASRIGQLYFGQYMRSSEARFLLESYVFYEAILNRGYFEGGKEHGGVSVRFKELRFYARFLLVALILNFRNVIDLLVDRFKALVDDSRATFR
1104	sp_215610_wuyg	-0.87426	-0.853419	#N/A	uncharacterized protein	MSYPSGYGTQYAPPSAPPPPGYPGYPPQQGYEGFHHNPPGPPPPGYGQYHHHQPGGYQGYFNDGYPPPPPPPQPYPPTQVYHCEHSHNHHNGCLSFLKG

PCT/IL2019/051000

	A	B	C	D	E	F
1		correlation coefficient with baits used in the analysis				
2	coexpressed gene	sp_107620_kqnh (*SOAP1*)	sp_107660_kiqg (*SOAP2*)	sp_107670_ptqx (*SOAP2-like*)	annotation	aa seq
1105	sp_215860_zffj	#N/A	-0.876382	#N/A	CBL-interacting protein kinase 1 (CIPK1); CONTAINS InterPro DOMAIN/s: NAF/FISL domain (InterPro:IPR018451), Serine/threonine-protein kinase-like domain (InterPro:IPR017442), Protein kinase-like domain (InterPro:IPR011009), Serine/threonine-protein kinase, active site (InterPro:IPR008271), NAF domain (InterPro:IPR004041), CBL-interacting protein kinase (InterPro:IPR020660), Protein kinase, catalytic domain (InterPro:IPR000719), Calcium/calmodulin-dependent protein kinase-like (InterPro:IPR020636); BEST Arabidopsis thaliana protein match is: CBL-interacting protein kinase 17 (TAIR:AT1G48260.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	GVFHRDLKLKNILVDRKGNIKVSDFGLSALPQHF G
1106	sp_216020_owkw	-0.809533	#N/A	-0.881788	unknown protein; Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	MLDEINWAKTRLVSDNSEQSAEQLKSEINNLKPE LSATDTNTLEEQSKWGNGLKLK
1107	sp_216350_xoui	#N/A	#N/A	-0.952214	SNARE-associated protein-related; Has 39 Blast hits to 39 proteins in 19 species: Archae - 0; Bacteria - 0; Metazoa - 6; Fungi - 0; Plants - 28; Viruses - 0; Other Eukaryotes - 5 (source: NCBI BLink).	MSNSDTISIGIAANDDSYSVIPTIFSPECDNNNVN DDNIAGSSEALAEGISTMLAGVIRDFDSKAEDAS RSQFQLISALDRLTS

[00960] Down regulation of SOAP3 and SOAP4 expression led to reduced accumulation of medicagenic acid derived saponins and caused accumulation of glycosylated intermediates (augustic acid, bayogenin and hederagenin; (**FIGURE 25**, **FIGURES 26A-26L**, **FIGURE 27**, **FIGURES 28A-28C**).

[00961] The structure of accrued pathway intermediates demonstrated that SOAP3 (SEQ ID NO: 52) is a C-2 hydroxylase and SOAP4 (SEQ ID NO: 54) is a C-23 oxidase (**FIGURE 20A**). CYP2 (CYP72A656; Sp_148230_dgra) (SEQ ID NO: 72 [gene] and SEQ ID NO: 73 [polypeptide]), which was also identified, displays high homology to SOAP4 (almost 92% at the amino acid level), likely exhibits the same activity in the saponin biosynthesis pathway. Additionally, reduced saponin content was observed in plants with silenced: SOAP6 (UGT74BB2; Sp_170930_hjgq; SEQ ID NO: 55 [gene] and SEQ ID NO: 56 [polypeptide]), SOAP7 (UGT79K1; Sp_020820_yeau; SEQ ID NO: 57 [gene] and SEQ ID NO: 58 [polypeptide]), SOAP8 (UGT79L2; Sp_113700_suxh SEQ ID NO: 59 [gene] and SEQ ID NO: 60 [polypeptide]) and SOAP9 (UGT73BS1; Sp_170320_dmqi SEQ ID NO: 61 [gene] and SEQ ID NO: 62 [polypeptide]) glycosyltransferases (**FIGURE 29**).

[00962] The same plants accumulated pathway intermediates lacking certain sugars: deoxyhexoses (SOAP6 (SEQ ID NO: 56) and SOAP7 (SEQ ID NO: 58)), hexose (SOAP8 (SEQ ID NO: 60)) and pentose (SOAP9 (SEQ ID NO: 62)) (**FIGURES 30A-30C**). Moreover, VIGS of one of the five co-expressed acyltransferases (AT) (i.e., SOAP10; Sp_125800_kzws; SEQ ID NO: 63 [gene], SEQ ID NO: 64 [polypeptide]) resulted in an increased ratio of desacetyl to acetylated saponins (from 20 to 80%) (**FIGURES 31A-31C**). The acetyltransferase activity of recombinant SOAP10 expressed in *E. coli* was subsequently confirmed *in vitro*, as it acetylated Yossoside IV and other Yossosides (e.g., Yossoside VI and Yossoside XII) with two deoxyhexoses attached at the C-28 position (**FIGURES 14A-14C**).

Example 18: Validation of SOAP Genes

[00963] *Objective*: To validate the function of all nine genes [i.e., β-amyrin synthase (bAS; SOAP1; SEQ ID NO: 45); cytochrome P450 (CYP450s; SOAP2, 3, 4; SEQ ID NO: 46, 51, 53, respectively); UDP-glycosyltransferases (UGTs; SOAP 6, 7, 8, 9; SEQ ID NO: 55, 57, 59, 61 respectively) and acyltransferase (AT; SOAP10; SEQ ID NO: 63)]

[00964] *Methods*: See Materials and Methods above.

[00965] *Results*: The nine genes of interest (SOAP 1-4 and 6-10) were transiently expressed in *N. benthamiana* leaves. Expression of the first four biosynthetic genes (SOAP1 to SOAP4) resulted in the formation of medicagenic acid (**FIGURE 20A**). Curiously, co-expression of all nine genes did not lead to production of any saponin that had previously been detected in spinach leaves.

Nevertheless, the accumulation of medicagenic acid and glycosylated derivatives [MA + hex; MA + 2 x hex, MA + 3 x hex) that are normally not present in spinach leaves were detected (**FIGURES 32A-32D**).

[00966] *Summary*: While reconstitution of medicagenic acid was achieved, expression of the nine genes SOAP1-4 and 6-10 did not result in the production of any saponins.

Example 19: Analysis of Cellulose Synthase Like G Spinach Homolog as a Component of the Saponin Biosynthetic Pathway

[00967] *Objective*: To identify the missing gene or genes necessary for reconstitution of the biosynthetic pathway of saponins.

[00968] *Methods*: See Materials and Methods above.

[00969] *Results*: At this point, it was realized that at least one enzyme was absent in order to reconstitute the saponin biosynthetic pathway. In an effort to identify the missing enzyme or enzymes the list of genes obtained by co-expression analysis was revisted. The only gene co-expressed with all baits in the co-expression set that appeared related to sugar metabolism was a Cellulose Synthase Like G spinach homolog (*SoCSLG*) (T. Richmond, Higher plant cellulose synthases, Genome Biology. 1, 3001.1–3001.6 (2000).). Although its contribution to saponin biosynthesis seemed unlikely based on this functional annotation, its function within this pathway was studied.

[00970] Notably, *SoCSLG* silencing resulted in high accumulation of medicagenic acid in spinach leaves (**FIGURES 33A-33F**; **FIGURES 34A and 34B**) suggesting its role in triterpenoid saponin biosynthesis. Next, *SoCSLG* (SEQ ID NOS: 65 OR 93) was transiently expressed together with SOAP1-4 (SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 51, SEQ ID NO: 53, respectively), in *N. benthamiana* leaves. The combined activity of all five enzymes resulted in the formation of medicagenic acid 3-O-glucuronide (MA-3-GlcA); the first glycosylated intermediate in the spinach triterpenoid saponin biosynthetic pathway (**FIGURES 20A-20C** and **FIGURES 32A-32F**; **FIGURES 35A-35C**).

[00971] Furthermore, combinatorial expression of CSLG (SOAP5, Sp_076690_ejcm (SEQ ID NOS: 65 or 93 [gene], SEQ ID NO: 66 [polypeptide]) with bAS (SOAP1; SEQ ID NO: 45) and CYP450s (SOAP2 [SEQ ID NO: 46], SOAP3 [SEQ ID NO: 51] and SOAP4 [SEQ ID NO: 53]) demonstrated that apart from medicagenic acid, SOAP5 [SEQ ID NO: 66] could glucuronidate other triterpenoid aglycones including oleanolic acid, augustic acid, hederagenin, gypsogenin and gypsogenic acid, though they were never detected in significant amounts in spinach (**FIGURES 36A-36F**). SOAP5 (*SoCSLG*) was identified as the spinach saponin glucuronic acid transferase. Conversely, inhibition of expression of SOAP5 (SEQ ID NOS: 65 or 93) could in some

embodiments, results in a decrease or inhibition of glucuronidating triterpenoid aglycones including but not limited to medicagenic acid, oleanolic acid, augustic acid, hederagenin, gypsogenin, and gypsogenic acid.

[00972] To examine if SOAP5 was indeed a missing component in the spinach saponin pathway a combination of all ten SOAP genes (See **Table 16**) were transiently expressed in *N. benthamiana* leaves. Metabolite analysis showed the presence of the most abundant spinach saponin Yossoside V signifying the identification of all required steps and enzymes in the native pathway (**FIGURES 33A-33F**).

[00973] *Summary*: Enzymes for all required steps in the native pathway of spinach saponin Yossoside V (**Compound 11**) have been identified and demonstrated to be functionally active in a heterologous plant system.

Example 20: Characteristics of the Triterpenoid Biosynthetic Pathway

[00974] *Objective*: To elucidate further details of the spinach saponin metabolic pathway.

[00975] *Methods*: See Materials and Methods above.

[00976] *Results*: The identification of SOAP5 (SEQ ID NO: 66) as the spinach saponin glucuronic acid transferase enabled deciphering the pathway further. It was discovered that SOAP6 (SEQ ID NO: 56) works directly on MA-3-GlcA linking a fucose to the carboxyl group at position C-28 of the aglycone, giving rise to Yossoside I (**Compound 7**). In fact, this makes SOAP6 the first characterized fucosyltransferase involved in triterpenoid metabolism (T. Louveau et al., Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family 1 provides insights into disease resistance and sugar donor specificity. The Pant Cell. 30, 3038 – 3057 (2018).) (**FIGURES 37A-37F**, **FIGURES 38A-38E**).

[00977] Expression of SOAP7 (SEQ ID NO: 57) generated the Yossoside II product only when expressed together with the other six SOAP1-6 genes (See **Table 16**). Based on the LC-MS/MS analysis, SOAP7 (SEQ ID NO: 58) is a glycosyltransferase transferring rhamnose onto the fucose of Yossoside I (**FIGURES 37A-37F**, **FIGURES 38A-38E**). Yossoside II is further converted into Yossoside III by SOAP8 (SEQ ID NO: 60), a glucosyltransferase that accepts as a substrate only an aglycone having a fucose and rhamnose attached to the C-28 position (**FIGURE 20A**; **FIGURES 37A-37F**, **FIGURES 38A-38E**). Expression of SOAP9 (SEQ ID NO: 61 [gene], SEQ ID NO: 62 [polypeptide]) showed that it is a xylosyltransferase attaching a pentose to glucuronic acid linked to the MA at C-3. In case of SOAP9, the sugar moieties decorating the triterpenoid backbone at C-28 are not crucial for its activity (**FIGURES 39A-39C**). The *in planta* and *in vitro* studies showed that SOAP10 takes part in last step of saponin biosynthesis in spinach transferring acetyl group to the

fucose moiety (**FIGURES 31A-31F**). The nucleotide and amino acid sequences of the acetyl transferases used for comparison are presented below in **Table 15**.

[00978] **Table 15: Nucleotide and Amino Acid Sequences of Acetyl Transferase Enzymes**

NAME	SEQ ID NO:	AA SEQ / NT SEQ
sp_074630_ygho_Spo04549_**AT1 gene**	**109**	ATGGCCAAATCTGAGCAAGAAACAATGGCCAAATCTGAGCAAGAAACATCCAT CAAACTTGTTTCAGAATGCTTTGTAAAACCAAAATATGAGATTAAATCCGCTAA GCAACCTTACCACTTAGGTCCCATGGATCTAGTTATGTTAACTATCGATCCTAT ACAAAAAGGTCTTGTCTTTACAATAAAGAATTCCCCACTCTTTTTGTCATCCGA ATCCCATGATAATATTGAAATTATTCGAACCAAAGTTGTGTCACGTTTATTAGA AAAGCTTAAACACTCACTTTCTATAGCTCTAGTCCACTTCTACCCGTTAGCAGG TCGTTTCACTACACAAAAACAACCCGAGCATAACACGAGCTTGGTCTTTATTGA TTGCAACAAAGGTCCCGGAGCGCGGTTCATCCACGCTACTTCCCTTGACTTTAC TATCTCCGATATACTTTCACCGGTTGATGTTTCCATCGTTCATTCTTTCTTTGATC TCGGTGAAAAGCATGTAAACTACGATTGTCATACTAAGGCGTTGCTATCGATCC AGGTAACAGAACTTTTAGATGGGGTGTTTATTGGGTTTAGCATGAGTCATAGTG TGGTTGATGGTACCTCTTTTATTCATTTTGTCAATACCTTGTCTGAAATTTTTAA ATCTGATGATTTTACCACTATTTCACGTGCCCCAATACTTAATTATAGGCCTTGT GATATTCCGATCCTTAAATTTCCGTTTCTTGATGTGGAGGGGTTTATATGTCGTG CGTATAACCCTGGGCCGTTAAGGGAAAGAATCTTCCACTTTTCACTAAATTCGA TGCTGAGACTCAAGGCCATGGCTAACCAAGAATGTGGTACCCAAAATGTTTTA TCATCTTTCCAAGCTTTGACTGCGGTTGTATGGAGGTCCATCACCCGAGTTCGG AACTTACCAAAGGATGAGCAAACCACGTGTTTTATGGCTATGGGTTCTCGAACT AGGCTCAACCCGCCTCTTTCGGATGACTATTTTGGGAATTTTATGATTAGTACC AAATTTGCTTGCAAGGCAGAGGAATTATTGGGTAACAGTTTAGGTTGGGTAGC AATGAATTTACGTAAAATCATTATGTCCACTGACGAGAAATCGATACTTGCTAC GTACAAAGCATTGGCTGATTCCCCAATAGTGATTCCGCGTGAAACGATCCCCG GTCCTCATGGGATGACCAGAGTAATAATTGGAGGATCTTCAAGGTTCGATATG TATGGGCCTGAATTTGGATTGGGTCGAGCTTTGGCCGCTCGCATGGGTTATGGG AATAAGGATGATGGGAAAATAACAGCAAATCCTGGGTGTGAAGGAGGTGGAA GTGTTGATTTGGAAATTTGCCTTAGGCCTCATATTATGGCCTCTCTTGAAGTTGA TCAAGAGTTTATGGGTTTTGTGTCCTAG
sp_123780_pgiy_Spo21561_**AT2 gene**	**110**	ATGACTCCAAATCTGCAAATAGTAACCAACGGAGGCAAACCGGAAAATGATG AAGCAGAACCCGTATCACCTACCGGACAATACTTCAACAGCAAAGTGTTGTCT GTTTGTGTCCTTGCCATTCTAGAAATTGATGTTCCTATAGATGACTCGTGTGTAA TTCCACAACTCCGTGATGTCTTCCTACCCATGAACCCCAGATTTTCATCTATCAT GATATCTGACAATAAAGATGTAAAACAATGGAAAAGAGTGGAAGTGAACCTTC AAGATCATGTTGTCGTCCCTAGCGTCCCAGATGGCTTATCGGTTGAATCATACG ACAAGTACTTTGATGAATATCTGACAAAAATAACAGTGGATCCATTACCACAG GATAGGCCTTTATGGGAACTTCATGTTATAAAATACCCAACAAGCAAAGCAGC GGGTCATTTCATCTGGAAGCTTCACCATGCACTTGGTGACGGCTACACTCTAAT GGGAGTACTTCTGTCCGGCGTGAACAGAGCAGATGATCCTTCCCTTCCGTTAAC TTTCCCTTCAACACGATCAAGCTCACTAGTTACAAACAACAAGATGAATATTAT CAGCTGGGTGCCAAGAACTTTTTCAGCAATCTACAACGGTGTTTATAATTTTGG ATGGAGTTTTCTAAAAAGCACTTGCAAGGCAGATGATAAGACACCTATCAGAT CCGGAAATGAAGGTCTGGGTTTCCACCCAATGAAGATCTCGACAATAGAACTA TCCCTAGACCAAATCAAATTTATCAAAACAAAACTCGGCGCAACGGTAAATGA CATTCTTGCAGGCATAATTTTCCTCGGGGTTCGAAAATACATGCAAGCAACTGA TACAGAATCTGGAAACTCAGAATCAACGGCATTGGTGCTGTTTAACACTAGGA ACATTGGAGGTTATATGACCGCTGAGCAAATGAAGAAAGCACAAATGAAAAT ATGGGGGAACCAATTTGCATTTTTGCATATAGCAATACCTCAATTAATCAATGA CAAATGCTCGAACCCCCTTGACTATGTCTATGAAGCACGAAAACAGATCTCTA GGTTCAAAAGCTCACCATCAGTCTATCTAACAGCTCAGTGCCTAGAGCTGCTAG GAAATGCAAAGGACCTGAGGCAGCAGCTGAATTTATCCAGAGTACAACGAATA AAGCAAGCATATTAA
sp_149	**111**	ATGTTAGAGCTAGCAGAAGACGAGGTGAAGAACTTCTTCAAGGTATGGGCAAT

180_nwmy_Spo15788_**AT3 gene**		AGTTTTTGCATCTTTAAGCTATTGTTATTACATAGGCAAGCTAATTAATCCAAA AGGTTATACAAGATTAGTAGCAATAATCCCAATTATTACTCTCTTTTTAGCACT TCCTTTAAATCTCACATCTTTTCATCTTGGTGGTATGACTTGTTTCTTTATTGCTT GGCTTGCTAATTTCAAACTCTTGCTTTTTGCTTTTGATAAGGGCCCACTTTGCGC TAATTCTTCAATCTCATTCGCCAAATTTCTTGCACTTTCTTGCTTACCCATCAAA ATCCAACACCCACCTCATAAAAAGTCATTAAAATCACACCCATCTATTTATAAT TACATCATTAAAGGGATACTTTTATGTCTAATAATTAAAATCTATGATTATGGT GATTACATTCATCCAAAAATCATATGGCTAATCTTTTTCTTCCACTCCTATTTTA CCATAGAGTTAGTCTTTGCATTCCTAGCAACATCGACTAATATTTTGTTAGGGC TCGAACTGGAGCCACAGTTCAATGAACCCTTAATATCAACCTCATTGCAAGACT TTTGGGGTAAGAGATGGAATATCATGGTGACAAGGATACTTAGGCCTACGGTG TACCTTCCCACACTAGAGTACTCCACTAAGGTCGTTGGACGCACGTGGGCCAC ACTTCCGGCGGTGATGTCCACGTTCTTTGTGTCAGCCATTATGCACGAGCTCAT CTTCTACTACTTGGGGCGCAACTGGCCCACATTCGAGGTGACGTGGTTCTTTCT CCTGCATGGATTATGTCTTTGTGTTGAGATTGTCGCTAAGAAGTTAGTTGGTGG GAAATGGAGGATCCCACGGTGGATTTCCGGCCCTGCCACGGTGTTGTTTGTGGT GGGTACTGGGTTTTGGCTGTTCTTGCCGCCGTTGTTGAAGGCTGGGTTGGATAC TAGACCGTTTCAAGAGTTTGCGGCCGTTGCCAAGTTTGTAAGGAGTTTGAAGGC AGCTCTCACATTTTG
sp_198340_focw_Spo13090_**AT4 gene**	**112**	ATGGCTCCTCCTTCTTCTTCTTCAACTACGGGTTCTGGTAATGGTTCTAGTTTTG CAGTCAATATAATGGCGTCGTTCTACATTTCCCCACAACAACCTTCCACCACAA ATTCACATTCTATCCCTCTCACTTTCTTTGACATTCCTTGGCTTCAATATCCTCC GCTCCAACCTCTCTTCTTCTTTCAACTTCCATCTACACCCCAATCTTCTTCTTCTT CTTCTTCTTCTTCTTCTTCTTCTTTCGACCACAACTTGTACTTGGAGTTTAGCTCC ACCATCCTCCCTAGGCTCAAACACTCCCTCGCTTCTGCCTTGCAATATTACTTTC CCTTTTCTGGAAAACTCACCACTACTACCCATACTATCCCGAATAACCTAGTTT TCTCGACAGACTCATCAGATTCTGTTGAGTTGACTGTTTCTCTGTGTGATGCTGA TTTTAATGGTCTATGCAGCTTTCTACCCAGGTCTACTCATCTCTTCCAACAATTG GTTCCCTCCTTGCCAAATATTGAATCCTCCAACCTCACTACATTCCCTGCACCTT TATTAGCTATTCAGATCACATTCTTTCCCACCTCTTCTCCTGGTTTCTCTATTGG CTTTGCTTCTCATCCTGTGCTTTCTGATCAGAGGACCTTCAGTAACTTCCTTTAC TCTTGGGCCTCTTTCTCCAAGTTTGATAATCTAAACATTTCACTTGCCCCTTCCT TCCCTGTCTCTGACAGGTCTGTCATTCTCGACCCTGATAGACTTGAGCCCCTTCT GTTGGAGCAGTGGTTGGGATTGGAGTCCAAACCAACCATGTCAACAAAGATGA AGCTACGTCCTCCTCCTGCTTATGTCCGTGGCTCGCTCCGGTCCACATTCGTCAT GGGCCCATCTGATATTGCTAATGCTACACAATGGTTACAAACCCAGTGTGAGA AGCTCAACAGATCATATCCTGTTCTCTTGTCACCCTACGTCGTCACTTGTGCCTT TATATGGACCTGTTTTCTGAGAGCCCGAGTCCAGAACAGTGCTGTTACTAAAGC CAAAGCAAAAGGCACCATGTACTTTGGATTTATTGCTGGTGGTATTACCCGTTT ACCCTATCGGGTACCTGCTAAGTATCTTGGCAACTGTGTCGGGTTTGGACGGGC AGCAGCGCAGAGGGAGGAGCTACTGAAGGAAGGTGAGGGGATGTTGGCAGCT GCTGATGCAATTGGGCTAACCATTAAAAAGTTGGATAAAGATGTTTTAGGAGG AGCTGAGAAATGGATATATGAATGGCAGACATTAATGGAATCCGAAGATCATA TTCATGTGGTTGGGTCGCCCAAGGTGAACCTTTATGAGACGGATTTTTGGTGGG GGAAACCGAAGAAGATAGAGGAAATTTCAACTGATGTTACCAGAGCCATCTCT CTTACACAGAGCAGGGACATGAAAAGGGGAATTGAAATTGGCCTCACTTTACC AAACTCCATTATGGATGACTTCTCCTCTATCTTCACTCAAGGCCTCCTTGTTTTT CAAAATTAG
sp_074630_ygho_Spo04549_**AT1 polypeptide**	**113**	MAKSEQETMAKSEQETSIKLVSECFVKPKYEIKSAKQPYHLGPMDLVMLTIDPIQK GLVFTIKNSPLFLSSESHDNIEIIRTKVVSRLLEKLKHSLSIALVHFYPLAGRFTTQKQ PEHNTSLVFIDCNKGPGARFIHATSLDFTISDILSPVDVSIVHSFFDLGEKHVNYDCH TKALLSIQVTELLDGVFIGFSMSHSVVDGTSFIHFVNTLSEIFKSDDFTTISRAPILNY RPCDIPILKFPFLDVEGFICRAYNPGPLRERIFHFSLNSMLRLKAMANQECGTQNVL SSFQALTAVVWRSITRVRNLPKDEQTTCFMAMGSRTRLNPPLSDDYFGNFMISTKF ACKAEELLGNSLGWVAMNLRKIIMSTDEKSILATYKALADSPIVIPRETIPGPHGMT RVIIGGSSRFDMYGPEFGLGRALAARMGYGNKDDGKITANPGCEGGGSVDLEICLR PHIMASLEVDQEFMGFVS
sp_123780_pgiy_Spo21561_	**114**	MTPNLQIVTNGGKPENDEAEPVSPTGQYFNSKVLSVCVLAILEIDVPIDDSCVIPQLR DVFLPMNPRFSSIMISDNKDVKQWKRVEVNLQDHVVVPSVPDGLSVESYDKYFDE YLTKITVDPLPQDRPLWELHVIKYPTSKAAGHFIWKLHHALGDGYTLMGVLLSGV NRADDPSLPLTFPSTRSSSLVTNNKMNIISWVPRTFSAIYNGVYNFGWSFLKSTCKA

AT2 polype ptide		DDKTPIRSGNEGLGFHPMKISTIELSLDQIKFIKTKLGATVNDILAGIIFLGVRKYMQ ATDTESGNSESTALVLFNTRNIGGYMTAEQMKKAQMKIWGNQFAFLHIAIPQLIND KCSNPLDYVYEARKQISRFKSSPSVYLTAQCLELLGNAKDLRQQLNLSRVQRIKQA Y
sp_149 180_nw my_Sp o15788 **_AT3 polype ptide**	**115**	MLELAEDEVKNFFKVWAIVFASLSYCYYIGKLINPKGYTRLVAIIPIITLFLALPLNL TSFHLGGMTCFFIAWLANFKLLLFAFDKGPLCANSSISFAKFLALSCLPIKIQHPPHK KSLKSHPSIYNYIIKGILLCLIIKIYDYGDYIHPKIIWLIFFFHSYFTIELVFAFLATSTNI LLGLELEPQFNEPLISTSLQDFWGKRWNIMVTRILRPTVYLPTLEYSTKVVGRTWA TLPAVMSTFFVSAIMHELIFYYLGRNWPTFEVTWFFLLHGLCLCVEIVAKKLVGGK WRIPRWISGPATVLFVVGTGFWLFLPPLLKAGLDTRPFQEFAAVAKFVRSLKAALT F
sp_198 340_fo cw_Spo 13090_ **AT4 polype ptide**	**116**	MAPPSSSSTTGSGNGSSFAVNIMASFYISPQQPSTTNSHSIPLTFFDIPWLQYPPLQPL FFFQLPSTPQSSSSSSSSSSSFDHNLYLEFSSTILPRLKHSLASALQYYFPFSGKLTTT THTIPNNLVFSTDSSDSVELTVSLCDADFNGLCSFLPRSTHLFQQLVPSLPNIESSNLT TFPAPLLAIQITFFPTSSPGFSIGFASHPVLSDQRTFSNFLYSWASFSKFDNLNISLAPS FPVSDRSVILDPDRLEPLLLEQWLGLESKPTMSTKMKLRPPPAYVRGSLRSTFVMG PSDIANATQWLQTQCEKLNRSYPVLLSPYVVTCAFIWTCFLRARVQNSAVTKAKA KGTMYFGFIAGGITRLPYRVPAKYLGNCVGFGRAAAQREELLKEGEGMLAAADAI GLTIKKLDKDVLGGAEKWIYEWQTLMESEDHIHVVGSPKVNLYETDFWWGKPKK IEEISTDVTRAISLTQSRDMKRGIEIGLTLPNSIMDDFSSIFTQGLLVFQN

[00979] To our knowledge, SOAP10 (SEQ ID NO: 64) is the first member of the benzylalcohol acetyl-, anthocyanin-O-hydroxy-cinnamoyl-, anthranilate-N-hydroxy-cinnamoyl/benzoyl-, deacetylvindoline acetyltransferase (BAHD) superfamily of acyltransferases reported to be involved in triterpenoid saponin biosynthesis.

[00980] **Table 16** below presents the amino acid and nucleotide sequences of the 10 enzymes, and genes encoding these enzymes, that comprise the biosynthetic triterpenoid saponin biosynthetic pathway producing Compound 11 (Yossoside V) in spinach.

[00981] **Table 16: Nucleotide and Amino Acid Sequences for Triterpenoid Biosynthetic Pathway Enzyme in Spinach**

NAME	ENZYME ACTIVITY	SEQ ID NO:	AA SEQ / NT SEQ
SOAP1	saponin β-amyrin synthase	45	atgtggaggttgaaggttggagaaggggctaatgacccatacttatatagcactaataactttgttg ggcgtcaaacttgggagtttgatcctaactatggcacccctgaggatattcaagaggtcgaagat gctcgccgcgatttttacaataatcggtttaaagtgaagccttgtaacgatctcttatggcgttttcag ttcttaagagagaaaaacttcaagcaaaccatacctcaagtgaaggtgggtgacggggaggag atcacatatgagaccgcctcgacgacattaaagagagcggtgaatattttcacagccttgcagtct gaacatggccattggccggctgaaattgctggccctcagttcttccttcctccttttggtattttgctta tacattacaggagatcttaactctgttttcggaccagaacatcgtagagaaattcttcgcagcatcta ctatcaccagaacgaagatggaggttggggattacatattgaaggacatagcaccatgttctgta ccgcactgaattacatatgtttacgaatgcttggaataggacctgatgaaggtgatgacaacgcgt gccctagagcgcgtaaatggattctcgaccatggtagcgttacacatatcccttcttggggtaaaa cttggttatctatactgggcctgttcgattggtctggaagtaacccaatgccacctgagttctggatc cttcctacttttctccctatgcatccagcaaaaatgtggtgctactgtcgaatggtgtatatgccaatg tcatacttgtatgggaagagattcgtaggtccaatcacacctctcattaaacaacttagggaagaa ctctacaacgaaccctttgaacaaattagttggaagaaaatgcgacatttgtgtgcaccggaggat ctctactatcctcatccattgattcaagacttgatgtgggacgctctttacctttttacggaacctctcc tgacccgttggcctttcaacaagttgatacgaaagaaagcattagaggttacaatggaacacata cattatgaagatgagaacagtcgttacataacaattggatgtgtcgagaaggttttatgtatgttagc

			ctgttgggtggaagaccctaaaggggatcattacaagaaacatcttgcaagagtacaagattaca tttggattgctgaagatggattgaaaatgcagagttttggaagtcaacaatgggattgtgggttttca gtacaggcattattagcttctaatcttagtctcgacgaaattggacctgctcttaagaaaggccattt cttcattaaggagtcacaggtgaaggacaatccatccggcgacttcaaagctatgcatcgccata tctcaaagggatcgtggactttctccgaccaagatcatggttggcaagtctccgattgcactgccg aaggccttaagtgttgtctaatcttatcaacaatgcccccggaaattgttggagaaaagatggacc ctgaacgcctttatgattctgtcaatgtcttgctttctctacagagtaataaaggagggctagctgcc tgggaaccagcaggggctcaagaatggttggaggtcctaaacccaacagaattctttgaagaca ttgtgattgaacatgagtatgtagagtgtacggcttcagcaattcaagctttaataatgttcaagaag ttatacccaggacacaggaaaaaagagattgaaaattttgtagtaaacgcagtcaagtaccttgaa aacacccaatatcctagtggaggatggtatggaaattgggggatttgtttcatatatggaacatggt ttgcactaggagggctagcagcaggtgggaagacatactataattgtgctgctgttaggaagggt gttgatttttttgcttactacacaaaaggaggatggtggttggggtgaaagttatatttcttgtcccaat aaggaatttgtgccaatagagggaaagtccaatttggttcagactggttgggctttgatgggtcta cttcatgctggacaggcggagagggatccaactcctctgcatcgtgcagcaaagcttttgattaat tcacaactcgaaaatggcgatttccctcaacaggaaataacaggagtcttcatgaagaattgcatg ttacattatccgatgtacagaagcatttatccactgtgggcaattgcagaatacagaaagcgtgttt cattaccttctatcaactctacttga
SOAP2	cytochrome P450	46	atggaactcttctttatgtgtgggctagtccttttcctctccctatctctagcctccttcttcctttttctata accaccatagaacccgggggtacaagctacccccgggcaagatggggtggccggtggtggg cgagtcatttgaatttttttcaaaccgggtggaaaggttacccggaaaagttcatatttgatagactg aacaagtacaccccaagccaagtgttcaagacttccatcgtaggagaaaaggttgcggttttatgt ggcgcggcgggtaacaagttcttgtactcaaacgagaacaagttagtacaagcttggtggccta gctctgttgataagatctttccttcttctacccaaacttcctccaaagaagaggctaagaagatgcg gaaactcctccctaacttcctcaagcccgaggctttacataggtacatacccatcatggatagcatt gccatccggcacatggagtccgggtgggagggaaaggacaaggtagaagtcttccctttggct aagaattacaccttctggctggcttgccgactcttcttaagcgtcgaggacccggctcatgtagcc aagttctccgaaccattcaacgacatagccgcagggatcatctcgatgccaatcgacctccccg gaacacccttcaaccgagggatcaagtcgtctaacgtcgtaaggaaagagttgagggccatcat aaagcagaggaaacttgacttagcagatggcaaggcttcacctacacaagatattctgtctcatat gttgttgacttgtactgaagatggcaagtttatgagtgaaatggatattgctgataagattctgggac ttcttattggtggacatgatactgctagtgcttcttgtacttttgttgttaagtttcttgctgagcttcctc acatatatgaaggtgtctacaaagagcaaatggagatagcaaattcaaaaaaagcaggagaact tctaaattgggaggacatacaaaaaatgaaatactcatggaatgtagcttgtgaagttatgcgtttg gctcctccacttcaaggtggtttcagggaagccctttctgatttcatgtataacggattccaaatccc caagggctggaagttatattggagtgcaaattcaacacatatgaacccggaatgcttcccggagc ccaagacgttcgacccatcgaggttcgacggtacgggaccagcaccatacacatacgtcccctt cggaggaggaccgagaatgtgcccgggcaaggagtatgcaaggctagagatattagtgttcat gcacaacgttgtcaagaggtttaaatgggaaaaaatgcttcctgatgagaaggttattgtcaatcc catgcctatcccagaacatggccttcctgtccgccttttccctcatcctcgaactgtagctgcttaa
SOAP2 -Like		47	atggagttcttctttctgtgtggtctagtcttttacttctccatatctctagcctccttcttcctttctataa ccaccatacaacccgggtttacccgctacccgcgggcgagatggggtggccggtggtgggcg actcgtttgaatttttttcaaaccgggtggaacggttacccggaaaatttcatctttgatagactcaac aaatacaccccaagccaagtgttcaagactttcatcctaagagaaaaggttgtgttttatttgagaa ggttgtag
SOAP1	saponin β-amyrin synthase	48	MELFFMCGLVLFLSLSLASFFLFYNHHRTRGYKLPPGKMGW PVVGESFEFFQTGWKGYPEKFIFDRLNKYTPSQVFKTSIVGEK VAVLCGAAGNKFLYSNENKLVQAWWPSSVDKIFPSSTQTSS KEEAKKMRKLLPNFLKPEALHRYIPIMDSIAIRHMESGWEGK DKVEVFPLAKNYTFWLACRLFLSVEDPAHVAKFSEPFNDIAA GIISMPIDLPGTPFNRGIKSSNVVRKELRAIIKQRKLDLADGKA SPTQDILSHMLLTCTEDGKFMSEMDIADKILGLLIGGHDTAS ASCTFVVKFLAELPHIYEGVYKEQMEIANSKKAGELLNWEDI QKMKYSWNVACEVMRLAPPLQGGFREALSDFMYNGFQIPK GWKLYWSANSTHMNPECFPEPKTFDPSRFDGTGPAPYTYVP FGGGPRMCPGKEYARLEILVFMHNVVKRFKWEKMLPDEKVI

			VNPMPIPEHGLPVRLFPHPRTVAA
SOAP2	cytochrome P450	49	MWRLKVGEGANDPYLYSTNNFVGRQTWEFDPNYGTPEDIQ EVEDARRDFYNNRFKVKPCNDLLWRFQFLREKNFKQTIPQV KVGDGEEITYETASTTLKRAVNIFTALQSEHGHWPAEIAGPQ FFLPPLVFCLYITGDLNSVFGPEHRREILRSIYYHQNEDGGWG LHIEGHSTMFCTALNYICLRMLGIGPDEGDDNACPRARKWIL DHGSVTHIPSWGKTWLSILGLFDWSGSNPMPPEFWILPTFLP MHPAKMWCYCRMVYMPMSYLYGKRFVGPITPLIKQLREEL YNEPFEQISWKKMRHLCAPEDLYYPHPLIQDLMWDALYLFT EPLLTRWPFNKLIRKKALEVTMEHIHYEDENSRYITIGCVEKV LCMLACWVEDPKGDHYKKHLARVQDYIWIAEDGLKMQSFG SQQWDCGFSVQALLASNLSLDEIGPALKKGHFFIKESQVKDN PSGDFKAMHRHISKGSWTFSDQDHGWQVSDCTAEGLKCCLI LSTMPPEIVGEKMDPERLYDSVNVLLSLQSNKGGLAAWEPA GAQEWLEVLNPTEFFEDIVIEHEYVECTASAIQALIMFKKLYP GHRKKEIENFVVNAVKYLENTQYPSGGWYGNWGICFIYGT WFALGGLAAGGKTYYNCAAVRKGVDFLLTTQKEDGGWGE SYISCPNKEFVPIEGKSNLVQTGWALMGLLHAGQAERDPTPL HRAAKLLINSQLENGDFPQQEITGVFMKNCMLHYPMYRSIYP LWAIAEYRKRVSLPSINST
SOAP2 -Like		50	MEFFFLCGLVFYFSISLASFFLFYNHHTTRVYPLPAGEMGWP VVGDSFEFFQTGWNGYPENFIFDRLNKYTPSQVFKTFILREK VVFYLRRL
SOAP3	cytochrome P450 (C-2 hydroxylase)	51	atgatagaaatcgggtatattgtaaaatgggtaatttgtttagtgattgttagatgggtatggaagatt gtgaattgggtttggtttacaccaaaaaggcttgagaagtttctaagaaaacaaggtttagatgga aattcatacagatttttgttgggtgatctcaaagatatgtctaaaatgcgtaaagaagctagacaaa aacctattccttttactcatgacttctttcatcgtatcttgcctttccacaatcaccatttcaataaatacg gggaaagcttcttttcatggatggggcctataccagttgtgaatgttgcagaacaagagcaagtaa agggtgtgttcactaggataaaagagtttcagaaggccaaattaaacccacttgttgcattgcttgt ccctggacttgtgagcgctgaaggtgataaatgggtcaagcacaggaagctcatcaacccggct tttcatatggaaaagcttaagcttatgcatccagcatttggcgccagtgttttggatatggtgaacaa gtgggagaagatagtatctaaaacaggttcctctgaagtggatgtgtggccgtttgtttccagcct gactgcagatgctatctctcgtgctgcttttggcagtagctatgatgaaggaagaaagatatttgag ttggttcttgaacaaactgaaatcaccctacgccttctgcaatcagtttatatccctggatggatgtat gtgccaacaaagaccaacaggaggatgaaaacagtaaactctgaaatacaaaatttattaaccg ggataatcgttaagagaaagaaggcaatggaggccggcgaagctgccaaggatgatttgttgg ggatattgttggagtccaactacaaagatactgaaaatgttctcagtaataagaaaaaactaagca tgactttccaggaattgattgatgagtgcaaactgttctacttagcagggcaagagtcgacctcgg tgttgctagcatggacaatgattctgttgggaaagcacacagagtggcaagcacgagcacgaga agaagtagttgcaacgtttggtaaaaacgaacctgattttgaaggcttaaaccatttgaagatagtg acaatgatactgaatgaggtgttgaggttgtaccctccagtgtgtacaatcacccgtaagaatttca accacgacgtacagcttggaaatctgacagtccctcgtggtgctatggttacgatgtcagcatatc gtattcaaagagatcctaaaatatggggtgatgatgcaaaagagtttaacccacagagattttcag aaggggttgcaaaggctacaaaggggaatattgcattctttccgtttggttgggggccgcgaattt gcatcggacagaactttgcacttattgaagctaaaatggcagtgtccatggttttacaacgcttttct tttgagctatcaccgtcttatactcatgctcctaccactatcctcactcttcaaccccaacaaggtgct catctcattatacataagctcagggactaa
SOAP3	cytochrome P450 (C-2 hydroxylase)	52	MIEIGYIVKWVICLVIVRWVWKIVNWVWFTPKRLEKFLRKQ GLDGNSYRFLLGDLKDMSKMRKEARQKPIPFTHDFFHRILPF HNHHFNKYGESFFSWMGPIPVVNVAEQEQVKGVFTRIKEFQ KAKLNPLVALLVPGLVSAEGDKWVKHRKLINPAFHMEKLK LMHPAFGASVLDMVNKWEKIVSKTGSSEVDVWPFVSSLTAD AISRAAFGSSYDEGRKIFELVLEQTEITLRLLQSVYIPGWMYV PTKTNRRMKTVNSEIQNLLTGIIVKRKKAMEAGEAAKDDLL GILLESNYKDTENVLSNKKKLSMTFQELIDECKLFYLAGQES TSVLLAWTMILLGKHTEWQARAREEVVATFGKNEPDFEGLN

			HLKIVTMILNEVLRLYPPVCTITRKNFNHDVQLGNLTVPRGA MVTMSAYRIQRDPKIWGDDAKEFNPQRFSEGVAKATKGNIA FFPFGWGPRICIGQNFALIEAKMAVSMVLQRFSFELSPSYTHA PTTILTLQPQQGAHLIIHKLRD
SOAP4	cytochrome P450 (C-23 oxidase)	53	atgatttcaaagagcgcgagtggtgttgcaattgcttcattgggttgcattatcatactgtattggatg tggaaattactaaaagggctatggttaacaccaaaaaagctagagaaatgtctaaaacaacaag gtcttgttggtaattcctacaaatttttgattggagatatgaaagaaagctcaaaattgcgtaacgaa gctttacaaaaacccattcctttcactcatgattattacaaccgtattcagcctttcatccatcagattc tcaacaattctggtgcaggcaaaaatatctacacatggttgggaccagtgccaacaatactaatta cacaacctgagttaataaaggatgctttcaataggatgaacaattttcagaaaccaagattaaatcc atatactcaaatgctttcaactggacttccgaactatgagggtcagaaatgggctaaacacagga agcttctcaaccctgcttttcaacttgataagctcaagcttatgatccatacttttgaaacctgtgttac ggatacactgaataagtgggagaagctagtttgtaaaacaggttcttcagaggttgatatatggcc atatttgacaactttaacgggagatggtattgctagagctgcatttggaagtagctttgaagatgga agaagaatattcgagcttctcacactgcagaaggatattgttattagtcttctcaaatattcttatattc caggatttaaatatatgccaataaagggtaaccggaagatgaaagaagcggacaatgaaataaa acctctgttgacgaatataattaatagaaggaggaaagcgatggaggccggagaagctcccaa agacgacttgttagggatgctacttgaatccaatgcaaacgaggctcgacaagttaatgaaaatg aaagtggtagtagcaagcgaaaatctgatctaacgatgagcttccctgagatgatcgatgcttgc aagcagttcttcttggctggtcaagagaccacctcagtggccctaacatggacaatgcttttgttag ccaagcaccaagattggcaaacacgagctcgacaagaagtacttgctacatttggaatgaatac cccagactttgatggcatacataatcgtcttaagattgtgacaatgatactctacgaggtgttaaggt tgtatccgccagtccctgcaacatcgcgaagggttcatgatcgtgaaacaaagctaggagatttg gtaataccacaaggggtaggagtttcattttccatacttcatgcacacttgaaccctgaaatttggg gtgatgatgccaaagaattcaagcctgatagatttgcagaagggattgcaaaagcaacaaaagg gaataactcttacttccccttggttggggacctaggatttgcattggccaaaacttcgcactagttg aggcgaaaatggcattgtgtatgattttgcagcgtttctctttcgatctctcgccttcatacatccatg ctccgactagtctcatatcccttcaacctcagcatggtgcccacattattttacatcgattttaa
SOAP4	cytochrome P450 (C-23 oxidase)	54	MISKSASGVAIASLGCIIILYWMWKLLKGLWLTPKKLEKCLK QQGLVGNSYKFLIGDMKESSKLRNEALQKPIPFTHDYYNRIQ PFIHQILNNSGAGKNIYTWLGPVPTILITQPELIKDAFNRMNNF QKPRLNPYTQMLSTGLPNYEGQKWAKHRKLLNPAFQLDKL KLMIHTFETCVTDTLNKWEKLVCKTGSSEVDIWPYLTTLTG DGIARAAFGSSFEDGRRIFELLTLQKDIVISLLKYSYIPGFKYM PIKGNRKMKEADNEIKPLLTNIINRRRKAMEAGEAPKDDLLG MLLESNANEARQVNENESGSSKRKSDLTMSFPEMIDACKQF FLAGQETTSVALTWTMLLLAKHQDWQTRARQEVLATFGMN TPDFDGIHNRLKIVTMILYEVLRLYPPVPATSRRVHDRETKLG DLVIPQGVGVSFSILHAHLNPEIWGDDAKEFKPDRFAEGIAK ATKGNNSYFPFGWGPRICIGQNFALVEAKMALCMILQRFSFD LSPSYIHAPTSLISLQPQHGAHIILHRF
SOAP6	Glycosyl transferases (UDP-glycosyltransferase; Fucosyl transferase)	55	atgacgggaaaaggaagaacgatggaggtgatcatgatgccatttcaccaccaaggtcacttaa ccccgatgctccaattcgcgaagcgcttcgcttggaaaggtgctggctcgatccggatcaccctc gccaccaccctctccaccgcccaaaatatgaccaattccaaaaacaacaacaacaataacgatt acgatttcctgacggtcgaaagcatctacgacgataccgatgattctcagctcaaattcatgggtc gtatgggaaagttcaagtccgaagcctcactccaacttggccgtctaatcactactaagagcatc gacaataataaatgtatgctcgtttatgatgcgtatctgccttgggcactggatgtgggcaaggac cataacatacaggctgcggctttcttcgtccaggcttgtgcgtatatggcatccttttaccctatgtttt tagaggaatttgggtcggatgatcaacatcctgttgttgcggctgctaaggctgaatctgttcctag tttgtcggttgagctgccgtcgcgggaggaaatggaacgatacgcgccgaaatgtgcacaatcc ccgagttctgatgataaacccaatactgttaagaaatcgcttcaccctgtctaccggatggtggttt catcaattacaaccttcatcttgctgatttcgtgctcatcaactcctttgatcaccttgaacatcagct ggatgtcttagcacatgaagcagtagggtgtttcataacccattgcggttggaactcgataattga ggcgaccaactttggggttccgatgttggggatgccacagttcatggaccagtttttggatgctca ttttatggagaaggtttggggtgttggaattagggctaaggctgatgagaaaaactttgttacttgtg acgaaatcaagtgcggtgtcaatgaaattatgtacggagataaggcaaatatgatcaaggagaat

			gcagcaaagtggaaagacttggctaaggaggcagttggtgaaggaggcagttcagataagaat atcgacgagatcattaactggcttgcgtcctcctaa
SOAP6	Glycosyl transferases (UDP-glycosyltransferase; Fucosyl transferase	56	MTGKGRTMEVIMMPFHHQGHLTPMLQFAKRFAWKGAGSIR ITLATTLSTAQNMTNSKNNNNNNDYDFLTVESIYDDTDDSQL KFMGRMGKFKSEASLQLGRLITTKSIDNNKCMLVYDAYLPW ALDVGKDHNIQAAAFFVQACAYMASFYPMFLEEFGSDDQHP VVAAAKAESVPSLSVELPSREEMERYAPKCAQSPSSDDKPNT VKKSLHPVYRMVVSSITTLHLADFVLINSFDHLEHQLDVLAH EAVGCFITHCGWNSIIEATNFGVPMLGMPQFMDQFLDAHFM EKVWGVGIRAKADEKNFVTCDEIKCGVNEIMYGDKANMIK ENAAKWKDLAKEAVGEGGSSDKNIDEIINWLASS
SOAP7	Glycosyl transferases (UDP-glycosyltransferase)	57	atgggtaaaacagtagcagcttcagagcagttacacatagtaatgatcccatggtttgcttatgga cacattcttccctattttgagctttcaaacaaacttgctgaaaaaggccataaaatcaccttggtagtt ccaaacaaagtcaaacttgatttagaaccaaagatccgtcatccttccttaatcagcttacatgcatt cactgtcccacacattgaacccttacctccggggactgagacatgttcagacgtccccattgaact tcagcaccaccttgctgttgccatggacagggcccggcctgaggtggagtccatcatatcagcc attgacgatccgaagccggatctgttgttctacgataacgcttactgggtgcctgagatagccaca aagctggggatgaagtctgtgttttaccagattgcatgtgctttaagtatcacccgcattaagcaaa ccccgagtgcgagtgcgagtgcgagtgcgagtgccaagctcttcactttacccaagtgggtgct gacgcctaaggtgttaggcgacgcaagagccaattatggagaagggatcacctattaccagag agtgaagaaggctctaagttcctgtgatgctatcgccttacgcacatgccgggagattgaagggg aatcctctgatatcctggctgcacaatataacaagccagtcttcttaacaggtccggtcctacccga ggttgaattcctcccccctttggacaattcctgggctgagtggctagccaagtttgggcctaagtc cgtggttttatgttgcttcggaagccagtacgtccctgacaaggctcaactacaggagatggccct tgcccttgaggatactggtcttcccttcttgatgtccgttaagccacccacagagtgcgccaccata gaggaggcgttgccagaagggttctcagagagagtcaaggaacgcggggtggttcatggtgg atgggtgcaacagctacaaatactagctcacccatcagtgggttgctttatttgtcattgtgggtac gggtcaatgtgggaggggttgttgagtgataaccagctagtcttattaccacagcttcctgaccag ttaatgatggctcaaatgttggcagaaaagctcaaggtgggtgtgatggtggacagagaagaag atgatgggtgggtttccaggaagaacttgtgccaagcagtcaagtctgtcatggatcctcattccg agtttgcagctttactcaagaacaaccatgctaacttcagagacaagttgctaaccaacggttttat ggctaattaccttgaagtttttgaccaggatttgaaacgctttcttactgcaaattaa
SOAP7	Glycosyl transferases (UDP-glycosyltransferase)	58	MGKTVAASEQLHIVMIPWFAYGHILPYFELSNKLAEKGHKIT LVVPNKVKLDLEPKIRHPSLISLHAFTVPHIEPLPPGTETCSDV PIELQHHLAVAMDRARPEVESIISAIDDPKPDLLFYDNAYWV PEIATKLGMKSVFYQIACALSITRIKQTPSASASASASAKLFTL PKWVLTPKVLGDARANYGEGITYYQRVKKALSSCDAIALRT CREIEGESSDILAAQYNKPVFLTGPVLPEVEFLPPLDNSWAE WLAKFGPKSVVLCCFGSQYVPDKAQLQEMALALEDTGLPFL MSVKPPTECATIEEALPEGFSERVKERGVVHGGWVQQLQILA HPSVGCFICHCGYGSMWEGLLSDNQLVLLPQLPDQLMMAQ MLAEKLKVGVMVDREEDDGWVSRKNLCQAVKSVMDPHSE FAALLKNNHANFRDKLLTNGFMANYLEVFDQDLKRFLTAN
SOAP8	Glycosyl transferases (UDP-glycosyltransferase)	59	atgggtggagagaaagagttgcggatagtgatgttcccatggcttgcctttggacattttatcccat accttcacctttcaaacaaacttgctgaaaaaggccacaaaatcaccttgttgcttcccaacaaag ctaggcttcagttggagtcacttaaccttcatccttctctcataactttccattcaattactgtcccacc cctcgaaactctcccttatggcactgaaacaactgcggatatctccctcgaccaacatggtgaact ctcgatttccatggaccgcactcggcccgaggtggagtctttcctatcaacccataagcccgacc tcgtcctctacgacatggcccattgggtacccgagattgctgctaaggtcgggattaagtcagttt catacaacgttgtatgtgctattgctgtatctcatgttagacctagcctccctcttccaaaaggaacg gcagcacatgtacccctgccattgtcgtctgtccctaagtggagtcttaatcagcacggttcatcaa caccatattttggggaagggataacgttacttgaacggtctgtaatctccctctcgtctgcggatgc aatagccatccgcacgtgcagggagattgaaggggtatattgtgaccgtgttgctgccacattca acaagcctgtccttgtcaccagccacgccttgcctgatcttgaactcgaactctctccgttggaga ctcgctgggccgagtggctagctaggttcgagccagggtcagtgatcttttgctgccttggtagtc agcatgtcttagacgcaccccaactgcaagagttggccctggggttggaaatgacaggactacc

			cttcttgatggctgtaaaaccccctgtggggtgtacctccttggaggaggtgcttccagaagggtt taatgatcgggttagcgggcgaggggtggttcacggtgggtgggtgcagcagcagcagataat ggcgcacccatcgttagggtgctttgtgacccttgtgggtcttcgtcgatgtgggaggggttagt gagtgaaagtcagttggtattactcccacaactggcagaccaaactctgtatgccaagttaatggc agatgagctcaaggtgggtgtgaaggtggagagagaagagaacgggtggatgacgaagcga agtctatgtgaagctatcaagagtgtgatggatgaagatagtgatataagtcatgtagttaggaaa aatcatgctaaatatagaagtatgttgattagccctggctttattagtggctacattgacaacttcatc aaggatttacaagcccttgttccttag
SOAP8	Glycosyl transferases (UDP-glycosyltransferase)	60	MGGEKELRIVMFPWLAFGHFIPYLHLSNKLAEKGHKITLLLP NKARLQLESLNLHPSLITFHSITVPPLETLPYGTETTADISLDQ HGELSISMDRTRPEVESFLSTHKPDLVLYDMAHWVPEIAAKV GIKSVSYNVVCAIAVSHVRPSLPLPKGTAAHVPLPLSSVPKW SLNQHGSSTPYFGEGITLLERSVISLSSADAIAIRTCREIEGVYC DRVAATFNKPVLVTSHALPDLELELSPLETRWAEWLARFEP GSVIFCCLGSQHVLDAPQLQELALGLEMTGLPFLMAVKPPV GCTSLEEVLPEGFNDRVSGRGVVHGGWVQQQQIMAHPSLG CFVTLCGSSSMWEGLVSESQLVLLPQLADQTLYAKLMADEL KVGVKVEREENGWMTKRSLCEAIKSVMDEDSDISHVVRKN HAKYRSMLISPGFISGYIDNFIKDLQALVP
SOAP9	Glycosyl transferases (UDP-glycosyltransferase; Xylosyl transferase)	61	atggagctttcaaaccctaccacaacccctaccttaaacgcaacccaacccttacgaggctatttc attccattaatcacagttccaagccacatttccaatcttgttgacattgctaaactcttctcatcacgg ggagtacatgtgactatcctcaccacccaccacacctccctccgcttcaaacaatccatacatgat tggggcttcaaaatcgacctccacatcgtcgacttcccgttcagggaagtcggcttaccggaagg agtggaaaattacagtgatgccacccctgagcaagcaagccagcttttccaggcctttatgatgct tcagaagcctatggaggatgccattcgggctgctaagcccgacttcatcgtttccgataggtattat cattggtctactgatcttgcacgtgagcttgctattccacggctcatctttcatgtcagatgttattttg cattgtgtgctgctgaggttgttgccaagtttgcccctcatgagaaggttgaatctgacactgacct ctttttccttcctgacctccctgataccatccacatgacccgtttgcagcttcccgaatggattcaga cccgaaacatgttcactgttctcaatgagagaatggacgaggccgatagggagtgctacggtgtt attgtaaacagctgctacgagttggagagagcttatgctgacttctaccgcagcaacttgggtcga cgtgcttggtgtatcggtccctatccggtacactgcgacaaggttggtaaaaagaaaggagatga cagcaaaaaacattcatgttttgaatggcttgataaaatgggagaaggagaagttatatacgtgag ttttggcactctgtcgtgtttcagccctgctcaaatctcagagctggctactgcactcgaaatgtctg gtcacccgtttatctgggtagtaaggaatggtgagaaattgttacctgatggatttgaagaaagaat tacagagcaggacaaaggggtgttaataaaagactgggcgccacaagtgaaaatacttgagca cccagctgtaggcggatttctgactcattgtggatggaactcaactgtagaaagtttagcagcagg tgtgccaatggtcacatggccgcttggtgccgagcaattcttcaatgaaaagttgattagtggagtt ttgaaggtgggggtcgaggtcgggtctgagaagtggagtaggggtatagtaccgaatactgata tgattgagaaggataaaatagaaagggcgattaaagagctgatgagtaaagaacccgaggccg aggaaaggaggcagaaggttaaggagttgagtaaggctccaagaaatgcggttgaagaaggt ggttcgtctcgtaacaatttaagtgatttgattgaaaaattacaacgtttaaaggcgaatgaaatatc agtgtccacaaactcagaataa
SOAP9	Glycosyl transferases (UDP-glycosyltransferase; Xylosyl transferase)	62	MELSNPTTTPTLNATQPLRGYFIPLITVPSHISNLVDIAKLFSSR GVHVTILTTHHTSLRFKQSIHDWGFKIDLHIVDFPFREVGLPE GVENYSDATPEQASQLFQAFMMLQKPMEDAIRAAKPDFIVS DRYYHWSTDLARELAIPRLIFHVRCYFALCAAEVVAKFAPHE KVESDTDLFFLPDLPDTIHMTRLQLPEWIQTRNMFTVLNERM DEADRECYGVIVNSCYELERAYADFYRSNLGRRAWCIGPYP VHCDKVGKKKGDDSKKHSCFEWLDKMGEGEVIYVSFGTLS CFSPAQISELATALEMSGHPFIWVVRNGEKLLPDGFEERITEQ DKGVLIKDWAPQVKILEHPAVGGFLTHCGWNSTVESLAAGV PMVTWPLGAEQFFNEKLISGVLKVGVEVGSEKWSRGIVPNT DMIEKDKIERAIKELMSKEPEAEERRQKVKELSKAPRNAVEE GGSSRNNLSDLIEKLQRLKANEISVSTNSE
SOAP 10	Acyl transferase	63	atgggggaagtcaaccatgaagaagtagaaattgaaataatatcaatagaaaccataaaaccatc atcactacttccaccaaaaactcctccaaaaaccatcacactttctcacctcgatcaagctgcccct

	(BAHD acyl transferase)		ttgtactactatcctttacttttatactacactaacactactactaccccaacatcacaaattcgag ttgacataacaagtaccctaaaaacttcacttagcaaaacacttgacaaattccaccctattgcagg tcgatgtgtggacgactctacaatttgttgcaaccaccaaggaataccattcattgaaaccaaagtt gactccaatatcttggatgtcatgaactcgcctgagaaaatgaagttgcttatcaagtttctccctca tgcagagtttcaagatgtgactcgaccagtctcggatttaaaccatttggcgtttcaagtcaatgtttt ccggtgtggtggggtgatcattggctcctatgtgctccacaagctccttgatggaatctctcttgga actttctttaaaaattggtcaaccattgctaatgatgagcgagttaaggacgacgacctagtacaac ctgactttgaagccactattaaggcgttccctccgcgtacagcaactccaatgcttcctcgtaatca acaacttccaaaggcggctgaaaaaccaaataataatccagtcaaagttcttgtgacaaagagct tcgtatttgacattgtttctttaaagaagatgatgttcatggctaagagtgaattggttcctaaaccca ccaaatttgagaccgtgacagggtttatttgggaacaaaccttatcaacattgcgtaattctggagtt gaagttgaacatacatcgcttataatacctgtaaacatccgcccaaggatgagtccgccactccc aagaggatccatgggtaacttgctcaagaatgcaaaggcacaggccaacaccagcagcagca atgggcttcaagaccttgttaaagaaatccattcatctttgtctcaaacaacccagaaaattaatact cctcctcctcctcctcctcctcctactactactgctacaacaatccattcatctttgtctcaaacaa cccagaaaattaatactcctcctcctactactacaacaatccattcatctttgtctcaaacaacccag aaaattaatactactactactacagcagaggttattttgactaaacggaaagttgacaatccagtta cacagaatcgagaaggaaactacctcttcaccagttggtgcaagattgggttggatgaggctga cttcgggttcggaaagcccgtttgggtaattcccaacgatgggagacccccttaaggtcaggaat atgattttccttactgattataggcatcccgaaacaggcgttgaaggaattgcagcatggattacgt tggaagagaaacaaatgcaatgtttaaagtcaaacccagaattccttgcttttgctactcctaattag
SOAP 10	Acyl transferase (BAHD acyl transferase)	64	MGEVNHEEVEIEIISIETIKPSSLLPPKTPPKTITLSHLDQAAPL YYYPLLLYYTNTTTTTPTSQIRVDITSTLKTSLSKTLDKFHPIA GRCVDDSTICCNHQGIPFIETKVDSNILDVMNSPEKMKLLIKF LPHAEFQDVTRPVSDLNHLAFQVNVFRCGGVIIGSYVLHKLL DGISLGTFFKNWSTIANDERVKDDDLVQPDFEATIKAFPPRT ATPMLPRNQQLPKAAEKPNNNPVKVLVTKSFVFDIVSLKKM MFMAKSELVPKPTKFETVTGFIWEQTLSTLRNSGVEVEHTSL IIPVNIRPRMSPPLPRGSMGNLLKNAKAQANTSSSNGLQDLV KEIHSSLSQTTQKINTPPPPPPPPPTTTATTIHSSLSQTTQKINTP PPTTTTIHSSLSQTTQKINTTTTTAEVILTKRKVDNPVTQNRE GNYLFTSWCKIGLDEADFGFGKPVWVIPNDGRPPKVRNMIFL TDYRHPETGVEGIAAWITLEEKQMQCLKSNPEFLAFATPN
SOAP5	Cellulose synthase Like G (glucuronic acid transferase)	65	atggcaacttctcacattcgcaatgtccaattaaccagagccattgttaaccgtctccacatcttcct ccattccgtagccatcttctcgctcttctactaccgtttcacttccttcttcaactccgacatctccata cttgcttactccttactcaccaccgccgaactcttcttaacctttctatgggcttttactcaggctttcc ggtggcgtcccgtaatgagggaagtctccgggtacgaatccatcaaacccgaacaactaccgg gtttggatgtcttcattgtcactgctgacccgacaaaggagccagttctggaggtgatgaactccg tgatatcatccatggctttggattatccggttgatagactggcggtttacttgtcggatgacggtggt tctccgttgtcgaaggaggcgattaagaaggcttatgagtttgctaagctttggattccttttgtaat aagtataatgttaagacaaggtgtcctcaggctttcttctcgcctcttgctgatggggaaaggcttg attggaattctgagtttatggctgatcaattggaactccagaccaaatatgaagcttttagaaactat gtggagaaagaaagtggagataacaccaaatgtactgcagttcatgatcgacctccttgcgttga gattatacatgacaacaaacagaacggagaaagtgatgtgaagatgcccccttctggtttatgtagc cagggaaaagagacctggtcgtcctcatcgtttcaaagctggagcccttaatgctcttcttcgagt atccagtttaatgagcaatgcaccttacttattggtgttggattgtgatatgtactgccatgatccaac ttctgctcgtcaatctatgtgcttccatcttgacacaaacatggcttcctctcttgcatatgtgcaatac cctcaaattttctataatgttagcaaaaatgacatctatgatggccaagccagatcagctcatatga cgaaatggaaaggcatggatggactcagaggcccggtcttgaatggaactgggtattatttgaa gcgaaaagcattatttggaaagcctaataacgaagatgaatacctcaacagtcaaccagaaaag gcctttggctcctccacaaaattaattgctgcactaagagagaactccaagcaaaatcttgccata aaggaattgacagaagatgagttgtaccaagaggctagaaatttggctacttgcacatatgaagc aaacacactatggggcagtgaggtaggatattcgtatgagtgcttgttggagagtacattcactgg atatatgttacattgcagaggatggaaatctgtgtatctttacccaaaaagaccatgcttcttgggat gcacaacgattgatatgaaggatgctacggttcaactaataaaatggacctcctcattacttggaat tgccctgtcgaagtctagccctctaactttggccatgtccagtatgtcaatcctgcaaagcatgtgtt

			acgcgtacatcacatttacaggccttttgcagctccattggttatatatggtgttgtccttccaataa gcctattgaagggctttcctattttccctaaggtatcggatccatggattttgccatttgtgttgatattt gtatcctcccatcttcaacatctatatgaggtcctggaaagtgacaaatcagcaacacaatggtgg aatgaggtgagaatttggatgatgaaatcagtgacagcctgtttgtttgggttgacggaagcgata atgaagaagattggagtacaaactgcaacattcagattaacaaataaggtagttgagaaggaaaa gatggataaatacgagaaggagaggtttgatttctcaggagcagctatgcttatggttcctcttaat attttggtggtactaaatatggtgtcattcattggtggactcatgagggtcataatcaacaacagttat gatcaaatgtttgcacaacttttcctctccttttttgtcctacttcttagctaccctgttgttaagggatg gttataa
SOAP5	Cellulose synthase Like G (glucuronic acid transferase)	66	MATSHIRNVQLTRAIVNRLHIFLHSVAIFSLFYYRFTSFFNSDI SILAYSLLTTAELFLTFLWAFTQAFRWRPVMREVSGYESIKPE QLPGLDVFIVTADPTKEPVLEVMNSVISSMALDYPVDRLAVY LSDDGGSPLSKEAIKKAYEFAKLWIPFCNKYNVKTRCPQAFF SPLADGERLDWNSEFMADQLELQTKYEAFRNYVEKESGDNT KCTAVHDRPPCVEIIHDNKQNGESDVKMPLLVYVAREKRPG RPHRFKAGALNALLRVSSLMSNAPYLLVLDCDMYCHDPTSA RQSMCFHLDTNMASSLAYVQYPQIFYNVSKNDIYDGQARSA HMTKWKGMDGLRGPVLNGTGYYLKRKALFGKPNNEDEYL NSQPEKAFGSSTKLIAALRENSKQNLAIKELTEDELYQEARN LATCTYEANTLWGSEVGYSYECLLESTFTGYMLHCRGWKS VYLYPKRPCFLGCTTIDMKDATVQLIKWTSSLLGIALSKSSPL TLAMSSMSILQSMCYAYITFTGLFAAPLVIYGVVLPISLLKGF PIFPKVSDPWILPFVLIFVSSHLQHLYEVLESDKSATQWWNEV RIWMMKSVTACLFGLTEAIMKKIGVQTATFRLTNKVVEKEK MDKYEKERFDFSGAAMLMVPLNILVVLNMVSFIGGLMRVII NNSYDQMFAQLFLSFFVLLLSYPVVKGWL
SOAP5	Cellulose synthase Like G (glucuronic acid transferase)	93	ATGGCAACTTCTCACATTCGCAATGTCCAATTAACCAGAG CCATTGTTAACCGTCTCCACATCTTCCTCCATTCCGTAGCC ATCTTCTCGCTCTTCTACTACCGTTTCACTTCCTTCTTCAAC TCCGACATCTCCATACTTGCTTACTCCTTACTCACCACCGC CGAACTCTTCTTAACCTTTCTATGGGCTTTTACTCAGGCTT TCCGGTGGCGTCCCGTAATGAGGGAAGTCTCCGGGTACGA ATCCATCAAACCCGAACAACTACCGGGTTTGGATGTCTTC ATTGTCACTGCTGACCCGACAAAGGAGCCAGTTCTGGAGG TGATGAACTCCGTGATATCATCCATGGCGTTGGATTATCC GGTTGATAGACTGGCGGTTTACTTGTCGGATGACGGTGGT TCTCCGTTGTCGAAGGAGGCGATTAAGAAGGCTTATGAGT TTGCTAAGCTTTGGATTCCTTTTTGTAATAAGTATAATGTT AAGACAAGGTGTCCTCAGGCTTTCTTCTCGCCTCTTGCTGA TGGGGAAAGGCTTGATTGGAATTCTGAGTTTATGGCTGAT CAATTGGAACTCCAGACCAAATATGAAGCTTTTAGAAACT ATGTGGAGAAAGAAAGTGGAGATAACACCAAATGTACTG CAGTTCATGATCGACCTCCTTGCGTTGAGATTATACATGA CAACAAACAGAACGGAGAAAGTGATGTGAAGATGCCCCT TCTGGTTTATGTAGCCAGGGAAAAGAGACCTGGTCGTCCT CATCGTTTCAAAGCTGGAGCCCTTAATGCTCTTCTTCGAGT ATCCAGTTTAATGAGCAATGCACCTTACTTATTGGTGTTGG ATTGTGATATGTACTGCCATGATCCAACTTCTGCTCGTCAA TCTATGTGCTTCCATCTTGACACAAACATGGCTTCCTCTCT TGCATATGTGCAATACCCTCAAATTTTCTATAATGTTAGCA AAAATGACATCTATGATGGCCAAGCCAGATCAGCTCATAT GACGAAATGGAAAGGCATGGATGGACTCAGAGGCCCGGT CTTGAATGGAACTGGGTATTATTTGAAGCGAAAAGCATTA TTTGGAAAGCCTAATAACGAAGATGAATACCTCAACAGTC AACCAGAAAAGGCCTTTGGCTCCTCCACAAAATTAATTGC TGCACTAAGAGAGAACTCCAAGCAAAATCTTGCCATAAA

			GGAATTGACAGAAGATGAGTTGTACCAAGAGGCTAGAAA TTTGGCTACTTGCACATATGAAGCAAACACACTATGGGGC AGTGAGGTAGGATATTCGTATGAGTGCTTGTTGGAGAGTA CATTCACTGGATATATGTTACATTGCAGAGGATGGAAATC TGTGTATCTTTACCCAAAAAGACCATGCTTCTTGGGATGC ACAACGATTGATATGAAGGATGCTACGGTTCAACTAATAA AATGGACCTCCTCATTACTTGGAATTGCCCTGTCGAAGTCT AGCCCTCTAACTTTGGCCATGTCCAGTATGTCAATCCTGCA AAGCATGTGTTACGCGTACATCACATTTACAGGCCTTTTTG CAGCTCCATTGGTTATATATGGTGTTGTCCTTCCAATAAGC CTATTGAAGGGCTTTCCTATTTTCCCTAAGGTATCGGATCC ATGGATTTTGCCATTTGTGTTGATATTTGTATCCTCCCATC TTCAACATCTATATGAGGTCCTGGAAAGTGACAAATCAGC AACACAATGGTGGAATGAGGTGAGAATTTGGATGATGAA ATCAGTGACAGCCTGTTTGTTTGGGTTGACGGAAGCGATA ATGAAGAAGATTGGAGTACAAACTGCAACATTCAGATTA ACAAATAAGGTAGTTGAGAAGGAAAAGATGGATAAATAC GAGAAGGAGAGGTTTGATTTCTCAGGAGCAGCTATGCTTA TGGTTCCTCTTAATATTTTGGTGGTACTAAATATGGTGTCA TTCATTGGTGGACTCATGAGGGTCATAATCAACAACAGTT ATGATCAAATGTTTGCACAACTTTTCCTCTCCTTTTTTGTC CTACTTCTTAGCTACCCTGTTGTTAAGGGATGGTTA

[00982] *Summary*: This study uncovered numerous surprising details. For example, SOAP6 may serve as a template in the homology-based search for fucosyltransferase decorating quillaic acid in the biosynthetic pathway of QS-21, a potent vaccine adjuvant form *Quillaja saponaria* (D.J. Marciani, Is fucose the answer to the immunomodulatory paradox of Quillaja saponins? Int Immunopharmacol. 29, 908-913 (2015).). Also, SOAP10, a member of the BAHD superfamily of acyltransferases is involved in triterpenoid saponin biosynthesis.

Example 21: Production of Triterpenoid Intermediates in Yeast

[00983] *Objective*: To produce of triterpenoid intermediates in a further heterologous system, in the case the yeast *Saccharomyces cerevisiae*.

[00984] *Methods*: See Materials and Methods above.

[00985] *Results*: To rule out the involvement of additional plant enzymes in the formation of MA-3-GlcA we expressed SOAP5 in *Saccharomyces cerevisiae* (**FIGURE 33D**). Since yeast do not produce amyrin-type triterpenoids nor UDP-glucuronate, expression of enzymes that generate these precursors was required in order to provide the substrates essential for SOAP5 activity (P. Arendt et al., An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. Metabolic Engineering. 40, 165–175 (2017); T. Oka, Y. Jigami, Reconstruction of de novopathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in Saccharomyces cerevisiae. FEBS Journal. 273, 2645–2657 (2006)).

[00986] Expression of SOAPs1, 2, 3, and 4 in yeast resulted in formation of medicagenic acid (**FIGURES 40A-40B**). However, adding SOAP5 was not enough to produce medicagenic acid 3-O-glucuronide (**FIGURES 33A-33F**). To provide activated sugar for SOAP5 the UDP-glucose 6-dehydrogenase 1 from spinach (SoUGD1, Sp_189830_psca) catalyzing the conversion of UDP-glucose to UDP-glucuronate, was expressed (SEQ ID NO: 74 [gene] and SEQ ID NO: 75 [polypeptide]).

[00987] ATGGTGAAGATTTGCTGCATTGGGGCTGGTTATGTAGGAGGCCCTACTATG
GCTGTTATAGCACTCAAGTGCCCAAAGATTGAAGTTGTAGTGGTTGATATATCTGTGT
CTCGGATCACTGCATGGAACAGCGAGCAGCTTCCAATCTATGAGCCGGGTCTAGATG
ATGTGGTTAAGGAATGCCGTGGAAGGAACCTTTTCTTCAGCACTGATGTAGAAAAGC
ATGTTGCTGAGGCTGATATTGTTTTTGTCTCTGTGAATACCCCTACCAAAACCACAGG
TCTTGGAGCAGGCAAAGCTGCTGATTTGACCTACTGGGAGAGTGCTGCCCGTATGATT
GCTGATGTTTCAAAGTCTGACAAAATCGTTGTTGAGAAATCAACTGTGCCAGTGAAA
ACTGCTGAGGCAATCGAAAAGATTCTGACGCACAACAGCAAGGGAATCAACTACCAG
ATCCTTTCAAATCCGGAGTTCCTTGCTGAAGGTACTGCTATTCAGGACCTTTTTCACCC
TGACAGGGTTCTCATCGGTGGCCGGGAAACCCCAGCAGGCCTCAAGGCAGTCCAAGC
ATTGAAGGATGTGTATGCTCAATGGGTTCCTGATGAACGGATCTTAACCACCAATCTT
TGGTCTGCTGAGCTCTCAAAGCTTGCTGCCAACGCCTTCTTAGCACAGAGGATTTCAT
CTGTCAATGCAATGTCAGCTCTTTGTGAGGCTACTGGAGCAGATGTTACCCAAGTCGC
ATATGCTGTTGGTAAGGACAGTAGGATTGGGCAAAAGTTTTTGAACGCTAGTGTTGGT
TTTGGAGGGTCTTGCTTCCAGAAAGACATTCTGAACTTGGTTTACATTTGTGAGTGCA
ACGGTCTCCCTGAGGTGGCCGAGTATTGGAAACAGGTAATCAAGGTGAATGATTATC
AGAAGAATCGTTTTGTGAATAGGGTTGTGGCCTCCATGTTCAACACTGTATCGAACAA
GAAGATTGCTGTTCTTGGATTTGCATTCAAAAAGGATACAGGGGATACTAGGGAGAC
ACCCGCCATAGATGTGTGCAAGGGTTTGTTGGGAGACAAGGCAAGGTTGAGCATCTA
CGATCCACAAGTCACTGAGGATCAGATTCAGCGAGATCTCACCATGAACAAGTTTGA
CTGGGACCACCCAATTCACCTCCAGCCCACAAGTCCCACAACTGTTAAGCAAGTGAG
TGTTGTTTGGGACGCTTATGAGGCCACCAAAGATGCTCATGCTGTGTGTATTCTGACT
GAGTGGGATGAATTTAAGAAACTTGATTACAAAAGGATTTTTGACAACATGCAGAAG
CCAGCTTTTGTGTTTGATGGAAGGAACATTGTGAATGCAGATGAGCTGAGGCAGATT
GGGTTCATTGTGTACTCAATTGGTAAACCTTTGGATTCATGGCTCAAGGACATGCCTG
CTGTGGCTTAA (SEQ ID NO: 74)

[00988] MVKICCIGAGYVGGPTMAVIALKCPKIEVVVVDISVSRITAWNSEQLPIYEPGLD
DVVKECRGRNLFFSTDVEKHVAEADIVFVSVNTPTKTTGLGAGKAADLTYWESAARMIA

DVSKSDKIVVEKSTVPVKTAEAIEKILTHNSKGINYQILSNPEFLAEGTAIQDLFHPDRVLIG
GRETPAGLKAVQALKDVYAQWVPDERILTTNLWSAELSKLAANAFLAQRISSVNAMSAL
CEATGADVTQVAYAVGKDSRIGQKFLNASVGFGGSCFQKDILNLVYICECNGLPEVAEY
WKQVIKVNDYQKNRFVNRVVASMFNTVSNKKIAVLGFAFKKDTGDTRETPAIDVCKGL
LGDKARLSIYDPQVTEDQIQRDLTMNKFDWDHPIHLQPTSPTTVKQVSVVWDAYEATKD
AHAVCILTEWDEFKKLDYKRIFDNMQKPAFVFDGRNIVNADELRQIGFIVYSIGKPLDSW
LKDMPAVA (SEQ ID NO: 75)

[00989] Consequently, only yeast cells expressing SOAP5 together with all other five genes were able to produce MA-3-GlcA (**FIGURES 33A-33F**; **FIGURES 40A-40B**). Similar to *N. benthamiana* assays, SOAP5 was capable of glucuronidation of triterpenoid intermediates (oleanolic acid, gypsogenic acid and bayogenin) in yeast cells (**FIGURES 41A-41F**).

[00990] *Summary*: Demonstrated that it was possible to produce and glucuronidate triterpenoid intermediates in a system heterologous to the source enzymes.

Example 22: Examination of Cellulose Synthase Like Enzymes in Additional Plant Species

[00991] *Objective*: To identify and analyze functional Cellulose Synthase Like G (CSLG) enzymes in non-spinach plant species.

[00992] *Methods*: See Materials and Methods above.

[00993] *Results*: Glucuronic acid attached to a saponin aglycone via the hydroxyl at position C-3 is a common feature for many saponins found in plants, especially in the *Caryophyllidae*, primitive *Rosidae* and *Asteridae* (M. Henry, Saponins and Phylogeny: Example of the "Gypsogenin group" Saponins. Phytochem Rev. 4, 89–94 (2005)). To examine if conjugation of glucuronic acid to the triterpenoid backbone by a CSLG enzyme is not limited to spinach and its family, phylogenetic analysis of CSLG proteins was performed in more than 70 plant species including mosses, gymnosperms, and flowering plants. It appeared that spinach SOAP5 is related to CSLG proteins from the *Caryophyllales*, *Malvales*, *Apiales,* and *Fabales* orders (**FIGURES 42A-42E**).

[00994] Saponin profiling of selected species, as well as published data, demonstrated the presence of glucuronide-oleanane-type saponins in soy (*Glycine max*), alfalfa (*Medicago sativa*), Lotus japonicus and licorice (*Glycyrrhiza uralensis*) as well as in the *Caryophyllales* species like beetroot (*Beta vulgaris*) and quinoa (*Chenopodium quinoa*) (**FIGURES 43A-43C**).

[00995] The closest homologues of the spinach SOAP5 were then cloned from all of the aforementioned species and expressed transiently in *N. benthamiana* together with SOAP1-4. MA-3-GlcA formation was observed for all tested CSLG enzymes (**FIGURES 42A-42E**). **Table 17** below provides the nucleic acid and amino acid sequences of these SOAP5 homologs.

[00996] **Table 17: SOAP5 Cellulose synthase Like G homologs (having glucuronic acid**

transferase activity)

Name	SEQ ID NO:	Nucleic Acid/Amino Acid Sequence
BvCSLG polypeptide from red beet _(XM_010673823.2_)	94	MSSLHICKVQTTRAILSRFHILFHSLAILALFYYRFTSFSTTKSGILPWTLLTTAEV VLGFVWALTQAFRWRPVLRDVAGWDSIKEEQLPGVDVFICTADPIKEPVLEVM NTVLSAMALDYPAEKLGVYLSDDGGSPLTREAIKEASKFAKVWLPFCSKYGIKT RCPQAFFSSFCDGERLDWNQDFKADELVLKSKYEAFKNYVEKASEDESKCTMA HDRSPCVEIIHDNKQNGEGEVKMPLLVYVSREKRPNRPHRFKAGALNALLRVS GVLSNGPYLLVLDCDMYCNDPTSARQSMCFHLDPKLAPSLAFVQYPQIFYNTSK NDIYDGQARSAYKTKWQGMDGIRGPVLTGTGYYLKRKALYGQPHNEDEFLIN QPEKAFGSSTKFIASVSSNSKQNMALKEMTRDDLLEEAKNLATCAYESNTEWG NKIGYSYECLLESTFTGYLLHCKGWISVYLYPKRPCFLGCTTIDMKDAMVQLM KWTSGLLGVGISKFSPLTYAFSRMSILQSMCYGYFTFSALFGVSFLIYGIVLPVCL LKGVPVFPKVSDPWIGVFVVVFASSLLQHLYEVLSSDDSIKTWWNEIRIWIIKSV TASLFGTMDAIMKKIGIQKASFRLTNKVVDKEKLEKYEKGKFDFQGAAVFMVP LIILVVLNMVSFVGGLRRAIINKNCDEMFGQLFLSFFLLVLSYPVLEGIVTKVRKG RD
BvCSLG_ gene from red beet (XM_010673823.2)	95	CATCACAGCCACATGGGAAACAAAAAGTCTTAACCTCGGCCATCTCGTGTG CCTCGTGCACATTCAACATTGATTTTAATTGTGTGCTTTCTCCAAAATGTACC ATACTATATCATCTTGTCAGAATTCCAAACTCCAACTATAACCACCATTGAA GTCATCTACACACAAACACACACACTCTTTCTCCCTAAAAATGTCTTCTCTC CACATTTGCAAAGTCCAAACAACAAGAGCAATACTTAGCCGTTTCCACATA CTCTTCCACTCCTTAGCCATCCTTGCTTTATTCTACTACCGTTTTACATCGTTC TCTACCACCAAATCAGGCATACTTCCATGGACCTTACTAACCACAGCAGAG GTGGTTCTAGGCTTTGTATGGGCGTTAACACAGGCCTTTCGATGGCGGCCTG TGTTGCGAGATGTAGCTGGATGGGATTCCATCAAGGAGGAACAACTGCCAG GGGTGGACGTGTTCATATGCACAGCTGATCCAATAAAGGAGCCGGTGTTAG AGGTGATGAACACGGTGCTTTCGGCGATGGCATTGGATTACCCGGCAGAGA AGTTGGGTGTTTATCTTTCGGATGATGGAGGTTCTCCCTTGACTAGGGAGGC TATTAAGGAGGCTTCTAAGTTTGCTAAGGTTTGGCTTCCTTTTTGTAGTAAGT ATGGTATCAAGACTAGGTGTCCTCAGGCTTTCTTCTCTTCTTTTTGTGATGGG GAAAGACTTGATTGGAATCAGGACTTTAAGGCTGATGAATTGGTGCTCAAG TCAAAATATGAAGCTTTTAAGAATTATGTGGAGAAAGCAAGTGAAGATGAA AGCAAATGCACCATGGCACATGATCGTTCCCCTTGCGTTGAGATTATACATG ACAACAAGCAAAATGGAGAAGGCGAAGTGAAAATGCCCCTTTTGGTCTACG TATCCAGGGAAAAGAGACCAAATCGTCCTCATCGTTTCAAAGCCGGAGCTC TTAATGCTCTTCTCAGAGTATCAGGTGTATTGAGCAACGGGCCTTACTTATT GGTGTTGGACTGTGATATGTACTGCAATGATCCAACTTCTGCTCGTCAATCT ATGTGCTTTCATCTTGACCCAAAATTGGCTCCTTCACTTGCATTTGTGCAATA CCCACAAATTTTCTACAACACCAGTAAAAATGATATCTATGATGGCCAAGC TAGATCCGCGTACAAGACAAAATGGCAAGGAATGGATGGTATTAGAGGAC CAGTCTTGACAGGAACAGGGTATTACTTGAAGAGGAAAGCATTGTATGGAC AACCTCATAACGAAGATGAATTTCTCATTAATCAACCAGAGAAGGCCTTCG GCTCCTCCACAAAATTCATTGCGTCAGTTAGTTCAAACTCCAAGCAAAATAT GGCCTTGAAGGAAATGACAAGAGACGACTTGTTAGAAGAGGCTAAAAATTT GGCTACTTGTGCATATGAATCAAACACTGAATGGGGTAACAAGATTGGATA TTCGTATGAGTGTTTGTTGGAGAGTACATTTACCGGATATCTCTTACATTGC AAAGGATGGATTTCCGTGTATCTTTACCCAAAAAGACCCTGCTTCTTAGGAT GCACGACGATTGACATGAAAGATGCCATGGTTCAACTAATGAAATGGACCT CTGGATTACTAGGAGTTGGCATATCAAAGTTTAGCCCTCTAACTTATGCCTT TTCGAGGATGTCTATATTACAAAGCATGTGCTACGGTTACTTCACATTTTCA GCCCTTTTCGGAGTTTCGTTCTTAATATATGGCATCGTCCTTCCAGTATGCCT ATTGAAGGGTGTTCCTGTTTTTCCTAAGGTATCGGATCCATGGATTGGAGTT TTCGTGGTAGTATTTGCATCCTCCCTCCTTCAACATTTATACGAGGTTCTCTC AAGTGACGATTCCATTAAAACATGGTGGAACGAGATCAGAATTTGGATCAT CAAATCGGTAACAGCTTCCTTATTTGGAACAATGGATGCAATAATGAAAAA GATCGGCATACAAAAGGCTAGTTTCCGATTAACTAACAAGGTTGTGGACAA GGAAAAGCTCGAAAAATATGAGAAGGGCAAGTTTGATTTCCAAGGAGCAG

		CTGTGTTCATGGTTCCTCTTATCATTTTAGTGGTACTAAATATGGTGTCATTT GTTGGCGGATTAAGAAGGGCAATAATCAACAAGAATTGTGATGAAATGTTT GGGCAACTTTTCCTCTCATTCTTTCTCTTAGTTCTTAGCTACCCCGTTTTAGA AGGGATAGTAACAAAAGTAAGAAAAGGACGTGATTGAGATGAATTTGCATT GTTTGGTAAAAGATCCAAACTTAGAGAAAGAGATTGCGTAGGAGATCAAAG GAAACAATGTGAGAGATTTACAGGCTTCATGAGGCTTAAGACCTCATTAAT TTTTGTGACAATTTACAAATTCTGTCTCTATATTTTGGTCAAGACGTATCATT TGAAAATTTCCATGGTTAGGTAGTTAGATTTCAATGTTCCACGTTTGTAAAT AAGAAGATAAAATAAGGAAATTTGTGATTTTAGCTTTACATTTCTATGAGAT AGTCCTTTGTTGTGATGAAAGTTGTGTTCTTAGTGAAAATAATAAAACGTGA CATCAAAATTTTGAGTATAT
CqCSLG Protein from quinoa _(XM_021 866098.1)_	**96**	MAATHICKVQTKRVIINRIHILFHSLAILALFYYRFSSFSNPHISLFPWVLL TIADLVFTFIWAMTQAFRWRPVLHDVSGYESINPRDLPKIDIFICTADPTK EPVLEVMNSVISSMALDYPPEKMAVYLSDDGGSPLTREAIKKAVEFAKV WIPFCNMYGIKTRCPDAFFSALGNDERLHRDQDFNAHESLLKSKYEAFK KYVEKESGDINKCTVVHDREPCIEIIHDSKQDGEAEVKMPLVVYVAREK RPGHPHRFKAGALNALLRVSGLLSNAPYLLVLDCDMYCHDPTSARQSM CFHLDPNMSPSLAFVQYPQIFYNTSKNDIYDGQARSAHTTKWQGMDGL RGPVLNGTGYYLKKKAIYGRPHNEDEYLINEPEKAFGSSTKFIASLKENS NQDLVLKEFTNDLLQEARNLATCTYEANSLWGVEVGFSYDCLLESSYT GYLLHCKGWRSVYLYPKRPCFLGCTTIDMKDAIVQLIKWTSGLLGVAM SKFSPLTYAMSRMSILQSMCYAYITCSGLLAVPLFIYGVVLPFCLLKGVP VFPKVSDPWMLGFVFVFVSSHVQHLFEVLASDHSVQQWWNEVRIWIM KAITACLFGSTEAIMKKIGIQKTTFRLTNKVVEKEKLDKYEKGKFDFSGA AMLMVPLIILTILNLVSFVGGLVRVINHNNYDDMFGQLFLSFYLLLLSYP TFEGIVTKVTDKLRKKE
CqCSLG Gene from quinoa_(X M_021866 098.1)_	**97**	AAAAGTGAAACTGGTGACTTAGTCTTGTGTCACCCGGGTTCCTCGAGCACAT ATTCTTCAAGATTTTGGTTTTTTGTACGGAGTATTTATATACACAAAAATTA GGAACCAAATAGGAAGACTCATATCATTTCAAAATGGCGGCAACACACATT TGCAAAGTCCAAACCAAAAGAGTCATTATCAACCGTATTCATATCCTCTTTC ACTCTTTAGCCATTCTTGCTCTCTTCTACTACCGTTTCTCGTCTTTCTCCAACC CTCATATCTCCCTTTTTCCATGGGTATTATTGACTATCGCCGACCTCGTTTTC ACCTTCATTTGGGCCATGACTCAGGCCTTCCGTTGGCGCCCCGTCTTGCACG ATGTGTCTGGCTATGAGTCCATCAATCCACGCGATCTTCCAAAGATCGATAT TTTTATATGCACCGCTGATCCCACCAAGGAGCCTGTGTTGGAAGTGATGAAC TCGGTGATATCATCCATGGCGCTCGATTATCCGCCTGAAAAAATGGCGGTGT ATTTGTCGGATGATGGTGGTTCTCCTTTGACTAGAGAGGCTATTAAGAAGGC TGTTGAATTTGCTAAGGTTTGGATTCCTTTTTGTAATATGTATGGTATTAAGA CTAGGTGTCCTGATGCTTTCTTCTCCGCTTTGGGTAATGATGAAAGACTTCA TCGTGATCAAGACTTTAACGCTCATGAATCACTCCTCAAGTCGAAATACGA AGCTTTTAAGAAATATGTGGAGAAAGAAAGCGGTGATATTAATAAATGCAC CGTTGTGCATGATCGTGAACCTTGCATTGAGATTATACATGACAGTAAACA GGATGGAGAAGCTGAAGTGAAAATGCCCCTTGTAGTTTATGTAGCCAGGGA AAAGAGACCAGGTCATCCTCATCGTTTCAAAGCTGGAGCCCTTAACGCTCTT CTCCGAGTATCAGGACTATTGAGCAATGCGCCTTACTTATTGGTGTTAGACT GTGATATGTACTGTCATGATCCAACCTCTGCTCGTCAATCTATGTGCTTCCAT CTTGACCCGAACATGTCTCCCTCTCTTGCCTTTGTTCAATACCCTCAAATTTT CTACAACACTAGTAAAAATGATATCTATGATGGTCAAGCCAGATCAGCTCA TACGACGAAATGGCAAGGCATGGATGGACTCAGAGGACCGGTCTTGAATGG AACTGGGTATTATCTGAAGAAGAAGGCGATATATGGAAGGCCCCATAATGA AGATGAATACCTCATCAATGAACCAGAAAAGGCTTTTGGTTCTTCCACAAA ATTCATTGCTTCACTTAAAGAAAACTCGAACCAGGATCTTGTCTTGAAGGAA TTCACAAACGATTTGTTACAAGAGGCTAGAAATTTGGCTACTTGCACTTATG AAGCAAACTCGCTATGGGGTGTTGAGGTAGGGTTTTCGTATGATTGCCTGTT GGAGAGTTCATACACTGGATATCTCTTACATTGTAAAGGATGGAGATCTGT GTATCTTTATCCCAAAAGACCGTGCTTCTTGGGATGCACGACAATTGACATG AAGGATGCTATTGTTCAATTAATAAAATGGACTTCCGGATTACTTGGAGTTG CCATGTCAAAGTTTAGCCCTCTTACTTATGCCATGTCCAGAATGTCTATATT GCAAAGCATGTGTTACGCGTACATCACGTGTTCAGGTCTTCTAGCAGTTCCA CTCTTTATATATGGTGTTGTTCTACCATTCTGCCTACTTAAGGGCGTTCCTGT

		TTTTCCTAAGGTATCGGATCCATGGATGTTGGGTTTCGTGTTTGTATTTGTAT CCTCCCATGTTCAACATCTATTCGAAGTGCTAGCAAGTGATCATTCAGTGCA ACAGTGGTGGAATGAGGTGAGAATCTGGATCATGAAAGCGATAACAGCCTG CTTGTTTGGATCAACTGAAGCAATAATGAAGAAGATTGGGATACAGAAAAC AACATTCAGATTAACAAATAAGGTTGTGGAGAAAGAGAAGTTGGATAAATA CGAGAAGGGAAAGTTCGATTTCTCAGGAGCAGCAATGCTAATGGTTCCTCT CATCATTTTGACTATACTAAATTTGGTGTCGTTCGTTGGGGGACTTGTAAGG GTGATCAACCACAACAACTATGATGATATGTTCGGGCAACTTTTCCTGTCAT TTTATCTCCTACTTCTTAGCTACCCTACTTTCGAAGGGATTGTTACAAAAGTT ACAGACAAACTTAGAAAGAAAGAATAAGGAGTGATTGAGTAACTGCCTAG TACAGTTTTCACTTCACTTCTCTAGATTAGTCCTTGTTTTTGTTTATGTTTATT AAGATCAGCAACACTTGTAGACGGTTGCAATAATGAGTTCAATACCGTTTG TTCTGTCCCTCTTGCAAGAACAAGTATATAAATACTTTTCATTAGCCGGTTG CATTGTTGGATTCATATAGATGAATATTTCAAATATTTCATCTTTTTGAACTT ACACACTAATGATTTATATCTGGAATTTTGAAA
MsCSLG Protein from Alfalfa_ (MSAD_2 99835)	**98**	MATFTFHKETVQPLLPLRRAYIIFHFTCVLFLFYYRISNLFISYPWFLMTIAEIILSF LWFFNQAFRWRLVNRSVMTEKLPPEEKLPGLDIFVCTIDPEKEPTVDVMNTVIS AIAMDYPSNKLSIYLSDDGGSPITLFGIKEAFEFAKVWVPFCKKYDVKSRCPKFF FTALGENERLHRPREFEEVRDQIKKRLNRVDPNSSQVENSKEHMPTKAKYEKM QKNIEKFGSNLKNLCMVTDRPSRIEIINDQKEMPLVVYVSREKRPSVPHRFKGG ALNTLLRVSGLISNGPYVLVVDCDMNCNDASSAKQSMCFFLDPETSKDVAFVQ FPQMFHNLSKKDIYDSQTRTAFTTKWKGMDGLRGPGLTGSGNYISRSALLFGSP NQKGDYLLDALYNFGKSNMYVESLKALRGQQTKKQNISRDVILQEACEVASCS YERNTNWGNEVGFSYAIKLESTVTGYLLHCRGWRSTYLYPKRPCFLGCAPTDM KEGLIQPIKWSSELLLLAISKYSPFTYGLSRLPTIHCLTFCYLVSTTQFATAYILYG FVPQICFLKGIPVYPKVTDPWFIVFTVLYLSSQIHHYIEVISTGGSSMIWWNEQRS GIVKSIGCVFAIIETAKKKFGLNKAKFTLSDKAIDKDKLKKYEQGKFNFDGAAL LMAPVIVLLTINIVCFFGGLWRLLNVRDFDEMFGQLFLIIYILALSHPIVEGIISMK RKSG
MsCSLG Gene from Alfalfa_ (MSAD_2 99835)	**99**	ATGGCAACCTTCACATTTCACAAAGAAACAGTTCAACCATTGTTACCTCTAA GAAGAGCTTACATAATCTTCCACTTCACATGTGTCTTGTTTCTCTTTTACTAC CGTATCAGCAATTTGTTTATTTCATATCCATGGTTTCTAATGACAATAGCTG AGATTATTCTATCATTTCTATGGTTTTTCAACCAAGCATTCCGTTGGAGGCTG GTGAATCGTTCAGTTATGACCGAGAAATTACCGCCGGAGGAGAAGTTGCCG GGACTCGACATATTTGTGTGTACCATTGATCCTGAAAAAGAACCAACGGTT GATGTTATGAACACTGTTATTTCTGCTATTGCAATGGATTACCCTTCTAATA AACTTTCTATTTATCTTTCTGATGATGGAGGTTCTCCTATTACTCTTTTTGGG ATCAAAGAGGCTTTTGAATTTGCTAAAGTTTGGGTTCCTTTTTGTAAAAAAT ATGATGTTAAGTCAAGGTGTCCTAAGTTTTTCTTCACTGCTTTGGGTGAGAA TGAACGACTTCATCGACCTCGTGAATTTGAAGAAGTGAGGGACCAGATTAA GAAGAGATTAAATAGAGTGGATCCTAACTCATCACAAGTTGAAAACTCCAA GGAACATATGCCCACCAAAGCCAAATACGAGAAAATGCAGAAAAATATTG AGAAATTCGGAAGCAACCTAAAGAATCTTTGTATGGTGACCGATAGACCTT CTCGGATCGAGATCATTAATGACCAAAAAGAAATGCCACTAGTTGTTTATG TATCTCGTGAAAAAAGACCATCTGTTCCTCACAGATTCAAAGGAGGAGCTC TCAATACATTGCTTAGGGTGTCAGGGCTAATCAGCAATGGACCTTATGTACT TGTCGTAGATTGTGATATGAATTGTAATGATGCATCATCAGCCAAACAATCC ATGTGCTTTTTTCTTGATCCTGAAACCTCTAAAGATGTTGCTTTTGTTCAATT CCCTCAAATGTTTCACAACCTTAGCAAGAAAGACATATATGATAGTCAGAC TAGGACTGCTTTTACGACAAAGTGGAAGGGAATGGATGGATTAAGAGGTCC AGGTCTAACTGGCAGTGGAAATTATATAAGTAGAAGTGCATTACTCTTTGG AAGTCCAAACCAAAAAGGGGACTATCTACTTGATGCTCTATACAACTTTGG CAAGTCTAACATGTATGTAGAATCACTAAAAGCGTTACGTGGTCAACAAAC TAAGAAGCAGAATATTTCAAGAGATGTAATTTTACAAGAAGCATGTGAAGT GGCTTCTTGTTCCTATGAGAGAAACACAAATTGGGGTAATGAGGTGGGATT CTCGTATGCTATAAAACTTGAGAGTACCGTTACTGGCTATCTCCTCCATTGT AGAGGATGGAGATCAACTTATCTTTACCCTAAAAGACCATGTTTCTTAGGAT GTGCTCCAACTGACATGAAAGAGGGATTGATTCAACCGATAAAGTGGTCAT CTGAACTTTTGTTGCTTGCAATCTCTAAATATAGCCCATTCACTTATGGCCTT TCAAGATTGCCCACTATTCATTGTTTAACTTTTTGTTACTTGGTAAGCACAAC CCAATTTGCAACAGCCTACATCTTATATGGATTCGTTCCTCAGATTTGCTTCT

		TGAAGGGAATACCTGTATATCCAAAGGTTACAGATCCTTGGTTTATAGTGTT TACAGTATTGTATCTATCCAGTCAAATTCATCATTATATTGAGGTAATTTCA ACTGGTGGCTCCTCGATGATTTGGTGGAATGAACAAAGAAGTGGGATTGTA AAATCAATTGGGTGCGTTTTCGCAATTATAGAAACAGCGAAAAAAAAGTTT GGGTTGAACAAGGCAAAATTCACTTTATCGGACAAAGCAATTGACAAAGAT AAGCTAAAGAAATATGAGCAGGGTAAGTTTAATTTTGATGGTGCAGCATTG CTCATGGCACCAGTGATTGTGTTACTCACAATAAATATTGTTTGCTTCTTTGG TGGTTTATGGAGACTACTCAATGTGAGGGATTTTGATGAAATGTTTGGTCAA CTTTTCCTCATTATCTATATACTTGCTCTAAGTCATCCTATTGTGGAGGGGAT TATATCTATGAAGCGGAAGAGTGGGTAG
GmCSLG Protein from Soybean_ (NM_001365113.1_soja)	**100**	MATFHTETVQSGLALSRLHILFHSVALLFLYYYRISHILLEPSFVWIFMTI AELIFGELWLFKQAFRWRPVSRAVMPEKLPSDGKLPALDIFVCTVDPEK EPTVQVMDTVISAIAMDYPSNKLAVYLSDDGGCPVTLYGIREASRFAKE WVPFCRKYGINSRCPKAFFSPMGEDERELLLLRNHEFLAEQEQLKAKYN IMQKNIDEFGRDPKNRSIVFDRPARIEIINEQSEIPLVVYVSRERRPNVPHT YKGGALNTLLRVSGLFSNGPYVLVVDCDMYCNDPSSAKQAMCFFLDPE TSKDIAFVQFPQMFHNLSMKDIYDSQHRHAFTTMWQGMDGLRGPGLS GSGNYLSRSALIFPSPYEKDGYEHNAQNKFGNSTMYIESLKAIQGQQTY KTSISRNVILQEAQAVASCSYEIDTNWGNEVGFSYVILLESTVTGYLLHC RGWRSTYLYPKRPCFLGCAPTDFMEGMLQLVKWSSELFLLGISKYSPFT YGISRIPILHNFTFCYFTSTCQYIVALIVYGIIPQVCFLKGTPVFPKVTEPW FVVFAILYVSSQSQHLIEVLYGGGSLGTWWDEQRIWIVKSIVGGIFGSILA IKKRFGLNKAKFILSNKVVAKEKFEKYEQGKFEFEDAALFMSPLVGLLIV NILCFFGGLWRLFNVKDFEKMSGQLFLLGYLAALSYPIFEGIITMKSKVQ
GmCSLG gene from Soybean_ (NM_001365113.1_soja)	**101**	TATATGCATGTTGACCGGTAAACATGGCGACCTTCCACACAGAAACC GTGCAATCAGGGTTGGCCTTGAGCAGACTCCACATCCTATTCCACTC GGTGGCACTCTTGTTTCTCTATTACTACCGCATAAGCCACATCTTACT GGAACCAAGCTTTGTATGGATTTTCATGACCATAGCGGAGCTTATCT TCGGCGAGCTCTGGCTCTTCAAACAGGCGTTCCGGTGGCGGCCCGTG TCGAGGGCCGTCATGCCGGAGAAGCTGCCGAGCGACGGCAAGCTTC CGGCGCTCGACATCTTCGTCTGCACGGTTGACCCCGAAAAGGAGCCG ACGGTGCAGGTGATGGACACCGTCATCTCCGCCATTGCCATGGACTA CCCCTCCAACAAGCTCGCCGTGTACCTTTCCGACGATGGCGGGTGTC CGGTGACTCTGTATGGGATCAGAGAGGCTTCTCGGTTCGCAAAGGAG TGGGTTCCGTTCTGCAGAAAGTATGGGATCAATTCACGGTGCCCCAA GGCCTTCTTCTCTCCCATGGGGGAGGATGAACGTGAACTGCTTCTTCT TCGCAACCATGAATTCTTGGCAGAGCAAGAACAACTCAAGGCTAAA TACAATATAATGCAAAAAAATATTGACGAATTTGGAAGAGACCCTA AAAATCGTTCCATTGTGTTTGATAGACCAGCTCGCATTGAGATTATA AATGAGCAATCCGAAATACCACTGGTTGTTTATGTGTCTCGTGAAAG AAGGCCAAATGTTCCTCATACATACAAAGGGGGAGCCCTCAACACA TTGCTCAGAGTCTCAGGGCTATTCAGTAACGGGCCCTATGTACTTGT AGTTGATTGTGATATGTATTGCAATGATCCATCATCAGCTAAACAAG CCATGTGCTTTTTTCTTGATCCTGAAACCTCCAAAGATATTGCTTTTG TCCAATTCCCTCAAATGTTTCACAACCTTAGCATGAAAGACATCTAC GATAGTCAACATAGGCATGCTTTTACAACAATGTGGCAAGGAATGG ATGGACTAAGAGGTCCAGGTCTTTCTGGTAGTGGCAATTACTTAAGT AGAAGTGCATTAATCTTTCCAAGCCCATATGAAAAAGACGGCTATGA ACATAATGCCCAAAACAAATTTGGCAACTCTACCATGTACATTGAAT CATTAAAGGCCATTCAAGGACAACAAACTTATAAAACGAGCATTTCA AGAAATGTGATTTTACAGGAAGCACAAGCAGTGGCCTCTTGTTCCTA TGAAATAGACACAAATTGGGGTAATGAGGTAGGATTCTCATATGTTA TATTACTGGAGAGTACAGTTACTGGCTATCTTCTTCACTGTAGAGGA TGGAGATCAACTTACCTTTACCCCAAAAGACCTTGTTTCTTGGGATGT GCCCCCACTGACTTCATGGAAGGCATGCTTCAGTTGGTGAAATGGAG TTCTGAACTTTTCTTGCTAGGAATATCCAAATACAGCCCTTTCACTTA TGGGATTTCAAGAATTCCTATTCTGCACAACTTTACCTTTTGCTACTT

		CACATCTACATGTCAATATATTGTTGCCTTAATAGTATATGGCATCAT TCCTCAAGTATGCTTCTTGAAAGGAACTCCTGTGTTTCCTAAGGTTAC AGAACCATGGTTTGTAGTTTTTGCAATATTATATGTATCCTCTCAAAG TCAACATTTGATTGAAGTCCTTTATGGTGGTGGCTCTTTGGGAACATG GTGGGATGAACAAAGAATATGGATTGTAAAGTCAATTGTTGGAGGC ATATTTGGATCTATACTAGCAATCAAGAAACGTTTTGGGTTAAACAA AGCAAAATTCATTTTATCAAATAAAGTTGTTGCCAAAGAGAAGTTTG AGAAATATGAACAAGGTAAGTTCGAGTTCGAAGATGCAGCTTTGTTC ATGTCTCCATTGGTTGGATTACTCATAGTGAATATTCTTTGCTTCTTT GGTGGTTTATGGAGACTATTTAATGTGAAAGATTTTGAAAAGATGTC TGGCCAACTTTTTCTACTTGGCTATCTGGCGGCGCTCAGTTATCCCAT TTTTGAGGGGATAATAACCATGAAAAGCAAGGTGCAATAGTAGTTTG TCAATGATTAGGCTAATTTAGGTATTTGAACTTTGTTCACAAATAATT TGCTTCATATGAAAATCTAAAGTGCATGCTAAATGTTTGTATCTTAAT ATGTAATTAGCGTGCTTTTATTTCATGCATGAGAATATGGCTATCGAT TTTAATTAGGAGCAAAATGTATGTTCTTACTCCATTTTTAATGCAATT TCTTTATTTTCTTGCCAATTAAA
LjCSLG Protein from Lotus japonicus _(Lj3g3v1 981230.1)_	**104**	MANFTLHTETVQAWLPLSRLHILIHSVFVILLLYYRTTRLIHAPTAPWIL MTVAEALLAVLWLFNQAFRWRPVSRSVKTEKLPRDENLPGLDIFVCTID PEKEPTAGVMDTVVSAVAMDYPPDKLSVYLSDDGGCAVTEYGIREACE FAKVWVPFCRKYGIKSRCPKVFFSPMGEDEEILRTDEFRAEQEKIKAQYE TMQKNIEKFGSDPKNCRIVTDRPSKIEVINEQSEIPRVVYVSRERRPSLPH KFKGGALNTLVRVSGLISNGPYVLAVDCDMYCNDPSSAKQAMCFFLDP ETSKYIAFVQFPQMFHNLSKKDIYDNQSRTAFKAMWQGMDGLSGPGLS GSGNYLSRSALLFGSPNQKGDYLLDAQNYFGESPLYIESLKAIRGQQTTK KNISRDESLLEAKVVASASYETNTEWGSEVGFSYGILLESTITGYLLHCR GWKSAYLYPKTPCFLGCAPTDIKEGMLQLVKWLSELCLFAVSKYSPFTY GFSRLPIMPTFTYCFLAASSLYAIVFILYGIVPQVCFLKGIPVFPKATDPWF AVFAVLYVATQIQHLIEVLSGNGSVSMWWDEQRIWILKSVTSVFAMIEG IKKWLGLNKKKFNLSNKAVDKEKVKKYEQGRFDFQGAALYMSPMVVL LLVNIVCFFGGLWRLFKEKDFADMFGQLFLLSYVMALSYPILEGIVTMK MKSG
LjCSLG Gene from Lotus japonicus _(Lj3g3v1 981230.1)_	**105**	ATGGCCAATTTCACTCTCCACACAGAAACCGTTCAAGCATGGCTCCC TCTAAGCAGACTCCACATTCTTATACACTCAGTGTTCGTCATCCTTCT CCTCTACTACCGCACAACGCGTCTCATCCACGCGCCGACCGCGCCGT GGATCCTGATGACCGTTGCGGAGGCTCTCCTCGCCGTGCTTTGGCTCT TCAACCAGGCCTTCCGGTGGCGACCGGTGAGCCGCTCCGTGAAGACA GAGAAGCTGCCGCGCGACGAGAATCTCCCCGGGCTGGACATATTTGT GTGCACGATTGATCCTGAGAAGGAGCCAACGGCAGGGGTGATGGAC ACGGTTGTTTCCGCCGTGGCGATGGATTACCCGCCGGATAAGCTATC CGTGTATCTTTCTGATGATGGTGGTTGCGCCGTGACGGAGTATGGGA TTAGAGAGGCTTGTGAGTTTGCCAAGGTGTGGGTTCCTTTTTGTAGA AAGTATGGGATCAAGTCGAGGTGTCCAAAAGTTTTCTTCTCTCCGAT GGGGGAAGATGAAGAGATTCTAAGGACAGATGAGTTCAGAGCAGAG CAAGAGAAGATCAAGGCCCAATACGAGACTATGCAGAAAAACATCG AGAAATTTGGTTCAGACCCCAAAAATTGTCGTATTGTGACTGACAGA CCCTCTAAGATCGAGGTTATAAATGAGCAATCAGAAATCCCACGTGT TGTGTACGTCTCTCGTGAAAGAAGGCCATCACTTCCTCACAAGTTCA AAGGAGGAGCTCTCAACACATTGGTCAGAGTGTCAGGTCTAATCAGC AATGGACCTTATGTGCTTGCAGTGGATTGTGATATGTATTGCAATGA TCCATCCTCTGCCAAGCAAGCAATGTGCTTCTTCCTTGATCCAGAAA CATCTAAATACATTGCATTTGTCCAATTCCCTCAAATGTTTCACAACC TTAGTAAGAAAGACATCTATGATAATCAATCTAGGACTGCTTTTAAG GCAATGTGGCAAGGCATGGATGGACTCAGTGGTCCAGGTCTTTCTGG CAGTGGTAACTACTTGAGTAGAAGTGCATTGCTATTTGGAAGTCCAA ACCAAAAAGGTGACTATCTGCTTGATGCTCAAAACTACTTTGGCGAG

		TCTCCCTTGTACATTGAATCATTGAAGGCCATCCGTGGACAACAAAC TACCAAAAAGAATATCTCAAGAGACGAAAGTTTACTAGAAGCTAAA GTGGTGGCCTCTGCTTCCTACGAGACAAACACAGAATGGGGCTCAGA GGTTGGATTCTCATATGGCATCTTACTGGAGAGTACTATTACTGGTTA CCTTTTGCACTGCAGAGGATGGAAATCAGCTTATCTTTACCCAAAAA CACCATGTTTCTTAGGGTGTGCCCCCACTGACATTAAAGAAGGCATG CTTCAGTTGGTGAAGTGGTTGTCTGAGCTTTGCTTGTTTGCTGTCTCT AAGTACAGCCCTTTTACATATGGGTTTTCAAGATTGCCCATTATGCCT ACCTTCACTTATTGTTTCCTGGCAGCTTCATCCCTATATGCTATTGTCT TCATCCTTTATGGCATTGTACCTCAAGTGTGCTTCTTGAAAGGAATCC CTGTGTTTCCAAAGGCCACAGACCCTTGGTTTGCAGTGTTTGCAGTAT TGTATGTAGCCACCCAGATTCAACATTTGATTGAAGTCCTTTCTGGCA ATGGCTCGGTCTCGATGTGGTGGGATGAACAAAGAATTTGGATTCTG AAGTCAGTTACTAGCGTATTTGCAATGATAGAGGGAATCAAGAAAT GGTTAGGATTGAACAAGAAAAAATTCAACCTGTCAAACAAAGCGGT TGACAAGGAGAAGGTCAAGAAATATGAGCAAGGTAGGTTTGATTTC CAAGGAGCAGCTCTGTACATGTCTCCAATGGTTGTGTTGCTCCTAGT GAACATTGTTTGCTTCTTTGGCGGTTTATGGAGACTGTTTAAGGAGA AAGATTTTGCAGATATGTTTGGTCAACTTTTCCTACTCAGCTATGTGA TGGCTCTCAGTTATCCCATTCTTGAGGGGATAGTAACTATGAAAATG AAGAGTGGGTAG

[00997] Silencing the beetroot CSLG (*BvSOAP5),* as in the case of spinach SOAP5, resulted in decreased levels of saponins and elevated accumulation of oleanolic acid (**FIGURES 44A-44C**). Similarly, suppressed expression of CLSG in alfalfa hairyroot resulted in decreased accumulation of saponins with attached glucuronic acid (**FIGURES 45A and 45B**).

[00998] *Summary*: These results demonstrate heterologous functionality of enzymes within the saponin biosynthetic pathway.

Example 23: Heterologous Production of Glycyrrhizin Using Cellulose Synthase like G from Licorice.

[00999] *Objective*: The discovery of CSLG glucuronosyltransferase activity fills a large knowledge gap in the biosynthetic pathways of numerous glucuronide-type triterpenoid saponins. One such pathway only partly deciphered to date generates glycyrrhizin; a triterpenoid saponin produced in licorice consisting of glycyrrhetinic acid decorated with two glucuronic acid moieties at the C-3 position. Glycyrrhizin (i.e. GL) and glycyrrhetinic acid monoglucuronide (i.e. GAMG), the single glucuronidated derivative, are important natural products long used in traditional Chinese and Japanese medicine and nowadays in the food, cosmetics and pharmaceutical industry. These molecules are low-calorie sweeteners with no glycemic index; GAMG is 941-fold sweeter than sucrose and 5-fold more than that of GL (K. Mizutani, T. Kuramoto, Y. Tamura, N. Ohtake, S. Doi, M. Nakaura, O. Tanaka, Sweetness of glycyrrhetic acid 3-O-beta-D-monoglucuronide and the related glycosides. *Biosci Biotechnol Biochem.* 58, 554-5 (1994)). Commercial production of GL and

GAMG depends on the currently limited availability of wild *G. uralensis* and inefficient industrial processing. Despite its importance, sustainable production of GL and GAMG in heterologous systems is currently impossible since the enzyme catalyzing attachment of the first glucuronic acid to the aglycone was not identified (Y. Nomura et al., Functional specialization of UDP-glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin. *Plant J.* doi: 10.1111/tpj.14409 (2019)). The goal here was to produce glycyrrhizin in a heterologous system

[001000] *Methods*: See Materials and Methods above.

[001001] *Results*: Some of the saponins, especially those produced by plants from *Fabales*, are of high importance due to their unique properties. Glycyrrhizin, triterpenoid saponin consisting of glycyrrhetinic acid decorated with two glucuronic acid moieties at position C-3 is produced by *G. uralensis* and widely used as calorie-free sweetener and a drug in many Asian countries. Four out of five enzymes involved in glycyrrhizin production were already characterized (bAS, CYP88D6 and CYP72A154, UGT73P12) but the enzyme responsible for attaching the first glucuronic acid to the aglycone was missing (Y. Nomura et al., Functional specialization of UDP-glycosyltransferase 73P12 in licorice to produce a sweet triterpenoid saponin, glycyrrhizin. Plant J. doi: 10.1111/tpj.14409 (2019).). The question asked here was if cellulose synthase like G from licorice (*GuCSL*) was able to perform this reaction.

[001002] The cellulose synthase like G gene from G. uraliensis was identified and used for expression in a heterologous system. The nucleotide and amino acid sequences of the expressed genes and the encoded polypeptides is provided below in **Table 18**.

[001003] **Table 18: Components of the triterpenoid Glycyrrhizin Biosynthetic Pathway in *G. uraliensis***

Name	Enzyme Activity	SEQ ID NO:	Nucleotide Sequence/Amino Acid Sequence
*GuCYP88D6_* GLYUR00056 1S00023451.1	β-amyrin 11-oxidase	76	ATGGAAGTACATTGGGTTTGCATGTCCGCTGCCACTTTGTT GGTATGCTACATTTTTGGAAGCAAGTTTGTGAGGAATTTGA ATGGGTGGTATTATGATGTAAAACTAAGAAGGAAAGAACA CCCACTACCCCCAGGTGACATGGGATGGCCTCTTATCGGCG ATCTATTGTCCTTCATCAAAGATTTCTCATCGGGTCACCCTG ATTCATTCATCAACAACCTTGTTCTCAAATATGGACGAAGT GGTATCTACAAGACTCACTTGTTTGGGAATCCAAGCATCAT TGTTTGYGAGCCTCAGATGTGTAGGCGAGTTCTCACTGATG ATGTGAACTTTAAGCTTGGTTATCCAAAATCTATCAAAGAG TTGGCACGATGTAGACCCATGATTGATGTCT CTAATGCGGAACATAGGCTTTTTCGACGCCTCATTACTTCC CCAATCGTGGGTCACAAGGCGCTAGCAATGTACCTAGAGC GTCTTGAGGAAATTGTGATCAATTCGTTGGAAGAATTGTCC AGCATGAAGCACCCCGTTGAGCTCTTGAAAGAGATGAAGA AGGTTTCCTTTAAAGCCATTGTCCACGTYTTCATGGGCTCTT CCAATCAGGACATCATTAAAAAAATTGGAAGTTCGTTTACT GATTTGTACAATGGCATGTTCTCTATCCCCATTAACGTACCT

			GGTTTTACATTCCACAAAGCACTCGAGGCACGTAAGAAGCT AGCCAAAATAGTTCAACCCGTTGTGGATGAAAGGCGGTTG ATGATAGAAAATGGTCCACAAGAAGGGAGCCA AAGAAAAGATCTTATTGATATTCTTTTGGAAGTCAAAGATG AGAATGGACGAAAATTGGAGGACGAGGATATTAGCGATTT ATTAATAGGGCTTTTGTTTGCTGGCCATGAAAGTACAGCAA CCAGTTTAATGTGGTCAATTACATATCTTACACAGCATCCC CATATCTTGAAAAAGGCTAAGGAAGAGCAGGAAGAAATAA CGAGGACAAGATTTTCCTCGCAGAAACAATTAAGTCTTAAG GAAATTAAGCAAATGGTTTATCTTTCTCAGGTAATTGATGA AACTTTACGATGTGCCAATATTGCCTTTGCAACTTTTCGAG AGGCAACTGCTGATGTGAACATCAATGGTTATATCATACCA AAGGGATGGAGAGTGCTAATTTGGGCAAGAGCC ATTCATATGGATTCTGAATATTACCCAAATCCAGAAGAATT TAATCCATCGAGATGGGATGATTACAATGCCAAAGCAGGA ACCTTCCTTCCTTTTGGAGCAGGAAGTAGACTTTGTCCTGG AGCCGACTTGGCGAAACTTGAAATTTCCATATTTCTTCATT ATTTCCTCCTTAATTACAGGTTGGAGAGAATAAATCCAGAA TGTCACGTTACCAGCTTACCAGTATCTAARCCCACAGACAA TTGTCTCGCTAAGGTGATAAAGGTCTCATGTGCTTAG
GuCYP88D6_ GLYUR00056 1S00023451.1	**β-amyrin 11- oxidase**	**77**	MEVHWVCMSAATLLVCYIFGSKFVRNLNGWYYDVKLRRKE HPLPPGDMGWPLIGDLLSFIKDFSSGHPDSFINNLVLKYGRSGI YKTHLFGNPSIIVXEPQMCRRVLTDDVNFKLGYPKSIKELARC RPMIDVSNAEHRLFRRLITSPIVGHKALAMYLERLEEIVINSLE ELSSMKHPVELLKEMKKVSFKAIVHVFMGSSNQDIIKKIGSSF TDLYNGMFSIPINVPGFTFHKALEARKKLAKIVQPVVDERRLM IENGPQEGSQRKDLIDILLEVKDENGRKLEDEDISDLLIGLLFA GHESTATSLMWSITYLTQHPHILKKAKEEQEEITRTRFSSQKQL SLKEIKQMVYLSQVIDETLRCANIAFATFREATADVNINGYIIP KGWRVLIWARAIHMDSEYYPNPEEFNPSRWDDYNAKAGTFL PFGAGSRLCPGADLAKLEISIFLHYFLLNYRLERINPECHVTSLP VSXPTDNCLAKVIKVSCA
GuCYP72A15 4_GLYUR000 890S0001907 1.1	**11-oxo-β- amyrin 30- oxidase**	**78**	ATGGATGCATCTTCCACACCAGGGGCTATCTGGGTTGTTCT GACAGTGATACTAGCTGCGATTCCCATATGGGCATGCCATA TGGTGAACACGCTGTGGCTGAGGCCAAAGAGGTTGGAAAG GCATCTCAGAGCTCAAGGTCTTCATGGTGACCCTTACAAGC TCTCACTTGACAACTCCAAGCAAACCTATATGCTCAAGTTG CAACAAGAAGCACAATCAAAATCCATTGGTCTCTCCAAAG ATGATGCTGCACCACGAATCTTCTCCCTTGCCCATCAAACT GTACACAAATATGGAAAGAACTCCTTTGCATGGGAAGGGA CAGCACCAAAGGTGATCATCACAGACCCAGAGCAAATTAA GGAAGTCTTTAACAAGATTCAGGACTTCCCCAAACCAAAAT TAAATCCCATCGCCAAGTATATTAGCATCGGTCTAATACAG TATGAGGGTGACAAATGGGCCAAACATCGAAAGATTATCA ATCCGGCATTCCACTTAGAAAAATTGAAAGGTATGCTGCCA GCATTTTCTCATAGCTGCCATGAAATGATTAGCAAATGGAA GGGGTTATTGTCATCAGATGGAACATGTGAGGTTGATGTTT GGCCCTTCCTTCAAAATCTCACTTGTGATGTAATTTCTAGG ACGGCATTCGGAAGCAGCTATGCAGAAGGAGCAAAAATAT TTGAACTTTTGAAAAGGCAGGGATATGCTTTGATGACAGCA CGATACGCACGCATTCCATTATGGTGGCTTCTACCATCAAC TACCAAAAGGAGGATGAAGGAAATTGAAAGAGGCATACGT GATTCACTTGAAGGTATCATTAGAAAACGAGAAAAAGCAT TGAAGAGTGGCAAAAGCACCGATGACGACTTATTAGGCAT ACTTTTGCAATCAAATCACATTGAAAATAAAGGAGATGAA AACAGTAAGAGTGCTGGAATGACCACCCAAGAAGTAATGG AGGAATGCAAACTTTTTTACCTGGCAGGGCAAGAGNTTGA AAATAAAGGAGATGAAAACAGTAAGAGTGCTGGAATGACC ACCCAAGAAGTAATGGAGGAATGCAAACTTTTTTACCTGGC AGGGCAAGAGACCACCGCAGCTTTGCTGGCCTGGACAATG

			GTGTTATTAGGCAAGCATCCTGAATGGCAAGCACGCGCAA GGCAGGAAGTTTTGCAAGTAACCATGATTTTATATGAGGTA CTCAGGCTGTACCCACCTGGGATTTACCTCACCCGAGCTCT TCGAAAGGATTTGAAACTTGGAAACCTTTTGCTACCTGCTG GAGTACAGGTTTCCGTACCAATACTTTTGATTCACCATGAT GAAGGTATATGGGGCAATGATGCAAAGGAGTTCAATCCTG AAAGGTTTGCTGAAGGAATTGCAAAGGCAACAAAAGGCCA AGTTTGCTATTTCCCTTTTGGATGGGGTCCTAGAATATGTGT TGGGCAAAACTTTGCCTTATTAGAAGCCAAGATTGTATTGT CATTGCTGCTGCAGAATTTCTCATTTGAGCTATCTCCGACTT ATGCACATGTTCCTACCACGGTGCTTACTTTGCAGCCAAAA CATGGGGCACCCATCATTCTGCATAAACTGTAA
GuCYP72A154_GLYUR000890S0001907 1.1	**11-oxo-β-amyrin 30-oxidase**	**79**	MDASSTPGAIWVVLTVILAAIPIWACHMVNTLWLRPKRLERH LRAQGLHGDPYKLSLDNSKQTYMLKLQQEAQSKSIGLSKDDA APRIFSLAHQTVHKYGKNSFAWEGTAPKVIITDPEQIKEVFNKI QDFPKPKLNPIAKYISIGLIQYEGDKWAKHRKIINPAFHLEKLK GMLPAFSHSCHEMISKWKGLLSSDGTCEVDVWPFLQNLTCDV ISRTAFGSSYAEGAKIFELLKRQGYALMTARYARIPLWWLLPS TTKRRMKEIERGIRDSLEGIIRKREKALKSGKSTDDDLLGILLQ SNHIENKGDENSKSAGMTTQEVMEECKLFYLAGQEXENKGD ENSKSAGMTTQEVMEECKLFYLAGQETTAALLAWTMVLLGK HPEWQARARQEVLQVTMILYEVLRLYPPGIYLTRALRKDLKL GNLLLPAGVQVSVPILLIHHDEGIWGNDAKEFNPERFAEGIAK ATKGQVCYFPFGWGPRICVGQNFALLEAKIVLSLLLQNFSFEL SPTYAHVPTTVLTLQPKHGAPIILHKL
GuCSL_GLYUR003152S00037491.1 (SOAP5 homolog; Cellulose Synthase Like G)	Cellulose Synthase Like G (Glucuronic acid transferase)	**80**	ATGGCAAGCTTCACCCTTCACACAGAAACCGTTCAGTCATG GCTACTCCTCAGCAGACTTCACATACTGCTGCACCTCGCAG TTGTACTGCTCCTCTTATACTACCGCATCACACGTTTCCCCT TCCATGCTCCSACTCTACCGTGGACTCTGATGACCGTAGGT GAGGCTATTATGGCAGTGCTGTGGTTCTTCAACCAGGCCTT CCGGTGGCGGCCGGTGAGCCGCTCGGTGATGACGGAGAAG CTGCCCAGCGACGCGAAGCTGCCGGGGCTTGACATATTCGT GTGCACGCTTGACCCCGAGAAGGAGCCCACCGTGGAGGTG ATGAACACTCTGGTCTCTGCCCTTGCCATGGACTACCCCCC TGACAAGCTCTCCGTTTACCTCTCCGACGATGGCGCCGCCC CGGTCACTCTTTACGGCGTGAGAGAGGCTTCTGAGTTCGCG AGGGTGTGGGTCCCTTTCTGCAAAAAGTATGGGATCAAGTC AAGGTGTCCCAAGGTTTTCTTCTCTCCCAGTGCTGAGGATG AACACCTTCTTCGCACCGACGAGTTCAGGTCAGAGCGAGA CCTCATCAAGGCTAAATACGAGAAAATGCAGAAAAATATT GAGAAATTTGGTTCGGATGCCAAAAATTGTCGTATGGTGAC TGACAGACCTCCTCGGATCGAGATATTGATTGACCAACCAG ACATGCCACGTGTTGTTTACGTGTCTCGGGAAAGAAGGCCA TCACTCCCTCACAAGTTCAAAGGAGGAGCCCTCAATACATT GCTCAGAGTCTCAGGTCTAATCAGCAATGGGCCTTATGTAC TTGTAGTGGACTGTGATATGTATTGCAATGACCCATCCTCA GCCAAACAAGCCATGTGTTTCTTTCTTGATCCTGAAACCTC TAAATYTATTGCATTTGTCCAATTCCCTCAAATGTTTCACAA CCTTGGCAAAAAAGACATCTATGACAATCAATCTAGGACT GCTTTTAAGACAATGTGGCAAGGGATGGATGGACTAAGAG GTCCTGGTCTTTCTGGCAGCGGTAATTACTTGAATAGAAGT GCATTACTATTTGGAAGTCCAAATCAAAAAGATGACTATCT GGATGATGCCCAAAACTACTTRGGCAAGTCTACCATGTACA TAGAATCACTAAAGGCCATTCGTGGACAAAAAACTATGAA AAAGAATATTTCAAGAGATGAAATTTTACGAGAAGCTCAA GTATTAGCCTCTTGTTCCTATGAGACAAACACAGAATGGGG AGCAGAGGTAGGATTCTCATATGGCATCTTACTGGAGAGTT

			CAATCACTGGCTATCTTTTYCACTGCAGAGGATGGAAATCA GCATATCTTTACCCAAAGACACCATGTTTCTTAGGGTGTGC CCCAACTGACATCAAGGAAGGAATGCTCCAATTGGTGAAG TGGTTGTCTGAATACTGCTTGCTRGGATTCTCTAAATACAG CCCTTTCACTTATGGCTTTTCAAGAATGCCCATTATGCCTAC CTTAGTCTATTGCTTCTTGACAACWACAACCCTTTATTCCA TTGTCTTCATCCTTTATGGCATTGTCCCCCAAGTTTGCTTCT TAAAAGGAATACCCGTGTTTCCAAAGGTCACAGACCCTTGG TTTGCAGTGTTTGCAACACTGTATATATCCACCCAGATTCA ACATTTGATAGAGGTCCTTTCTGGTGATGGCTCTGTGGCAA TGTGGTGGGATGAACAGKGAATCTGGATTCTGAAGTCAGT CACTAGCGTGTTCGCAATCATAGAGGCAGCTAAGAAAGGG TTAGGATTGAACAAGAAGAAATTCATGTTGTCAAACAAAG CAATTGACAAGGAGAAGCTCAAGAAGTATGAGCAAGGTAG GTTTGATTTCCAAGGTGCAGCTCTGTTCATGTCCCCAATGG TTGTGTTGCTCATAGTGAACGTTGTTTCCTTCATTGGTGGCA TATGGAGACTATTCAATGCAAAGGATATTGAAGATATGTTT GGTCAGCTTTTCCTAGTTAGTTATGTAATGGCCCTTAGTTAT CCCATTTTTGAAGGGATAATAACCATGAAAAGCAAGAGTG GATAG
GuCSLG_GL YUR003152S 00037491.1 (SOAP5 homolog; Cellulose Synthase Like G)	Cellulose Synthase Like G (Glucuron ic acid transferas e)	**81**	MASFTLHTETVQSWLLLSRLHILLHLAVVLLLLYYRITR FPFHAPTLPWTLMTVGEAIMAVLWFFNQAFRWRPVSRS VMTEKLPSDAKLPGLDIFVCTLDPEKEPTVEVMNTLVSA LAMDYPPDKLSVYLSDDGAAPVTLYGVREASEFARVW VPFCKKYGIKSRCPKVFFSPSAEDEHLLRTDEFRSERDLI KAKYEKMQKNIEKFGSDAKNCRMVTDRPPRIEILIDQPD MPRVVYVSRERRPSLPHKFKGGALNTLLRVSGLISNGPY VLVVDCDMYCNDPSSAKQAMCFFLDPETSKXIAFVQFP QMFHNLGKKDIYDNQSRTAFKTMWQGMDGLRGPGLS GSGNYLNRSALLFGSPNQKDDYLDDAQNYXGKSTMYIE SLKAIRGQKTMKKNISRDEILREAQVLASCSYETNTEWG AEVGFSYGILLESSITGYLXHCRGWKSAYLYPKTPCFLG CAPTDIKEGMLQLVKWLSEYCLLGFSKYSPFTYGFSRMP IMPTLVYCFLTTTTLYSIVFILYGIVPQVCFLKGIPVFPKV TDPWFAVFATLYISTQIQHLIEVLSGDGSVAMWWDEQX IWILKSVTSVFAIIEAAKKGLGLNKKKFMLSNKAIDKEK LKKYEQGRFDFQGAALFMSPMVVLLIVNVVSFIGGIWR LFNAKDIEDMFGQLFLVSYVMALSYPIFEGIITMKSKSG
GuUGAT_KT 759000.1	**Glycyrrh etinic acid glucuron osyltransf erase**	**82**	ATGACCATGGGTAACGAGAATCGGGAGCTGCACATAATCT TCTTCCCCTTTCTGGCGAACGGCCACATCATCCCCTGCGTG GACTTGGCCAGAGTCTTCGCCGCAAGAGGAATCAGAGCCA CCATAGTCACCACCCACCTCAACGTTCCCTACATTTCCAGA ACCATCGGAAAAGCCAACATCAACATCAGAACCATCAAGT TCCCTTCCACCGAAGACTCTGGCCTTCCCGAAGGCTGCGAG AATACCGAGTCAGCACTCGCCCCTGACAAGTTCATCAAGTT CATGAAGGCCACCCTGCTCCTGAGGGACCCACTTGAACAC GTGTTACAGGAAGAGCAACCACACTCTTGGTCGCCGACAT GTTCTTCCCTTGGGCCACCGACTCCGCCGCAAAATTCGGCA TCCCTAGGATCGTGTTCCACGGCCTCGGTTACTTCCCACTCT GCGTTCTTGCATGCACGAGACAGTACAAGCCTCAGGACAA GGTTTCATCTTACACGGAACCCTTCGTGGTTCCGAATCTCC CGGGTGAAATAACACTGACGAAGATGCAGCTGCCGCAGTT GCCTCAGCACGACAAGGTCTTCACCCAGTTGTTGGAAGAGT CAAACGAATCGGAGTTGAAGAGCTTCGGTGTGATTGTAAA

			CAGCTTCTACGAACTTGAACCGGTTTACGCGGATCATTACA GGAACGAGCTTGGGAGAAGAGCTTGGCATTTGGGTCCGGT TTCATTATGCAGTAGGGACACGGAGGAAAAATCGCGGAGG GGAAGGGAAGCTGCAATTGATGAGAACGAGTGCTTGAAGT GGCTTCAATCAAAGGAACCCAATTCGGTTGTTTATGTTTGT TTCGGTAGCATGATGGTTTTCAGTGACGCTCAGCTAAAAGA GATTGCGATGGGTCTTGAGGCTTCAGGGAAGCCATTCATAT GGGTGGTGAAGAAAGGAGGGGCTAAAAGTGAAGGTGAGA AATTGGAGTGGCTTCCAGAAGGGTTTGAGGAGAGAATGGG GGAAAGTAATAAGGGACTAATCATAAGGGGTTGGGCACCA CAGGTGATGATTTTGGACCATGGAGCGGTTGGAGGGTTTGT GACACATTGTGGGTGGAATTCAACGCTGGAAGGAGTGTGT GCAGGGGTGCCAATGGTGACTTGGCCCATGTATGGGGAAC AATTTTACAACGCCAAGTTTCTGACGGACATAGTGAAAATT GGGGTGGGTGTTGGGGTTCAAACGTGGATTGGGATGGGAG GAGGAGAGCCTGTGAAGAAGGAAGTGATAGAGCAGGCAG TGAGAAGGATAATGGTGGGGCAGGAAGCAGAGGAAATGA GAAACAGAGCCAAGGAACTGAGCCAGATGGCAAAGCGTGC TGTGGAGGAAGGAGGATCGTCTCACAACGATTTTAACTCTT TAATTGAGGATTTGAGGTCGCGTGCCCATTAA
GuUGAT_KT759000.1	**Glycyrrhetinic acid glucuronosyltransferase**	**83**	MTMGNENRELHIIFFPFLANGHIIPCVDLARVFAARGIRATIVT THLNVPYISRTIGKANINIRTIKFPSTEDSGLPEGCENTESALAP DKFIKFMKATLLLRDPLEHVLQEEQPHCLVADMFFPWATDSA AKFGIPRIVFHGLGYFPLCVLACTRQYKPQDKVSSYTEPFVVP NLPGEITLTKMQLPQLPQHDKVFTQLLEESNESELKSFGVIVNS FYELEPVYADHYRNELGRRAWHLGPVSLCSRDTEEKSRRGRE AAIDENECLKWLQSKEPNSVVYVCFGSMMVFSDAQLKEIAM GLEASGKPFIWVVKKGGAKSEGEKLEWLPEGFEERMGESNK GLIIRGWAPQVMILDHGAVGGFVTHCGWNSTLEGVCAGVPM VTWPMYGEQFYNAKFLTDIVKIGVGVGVQTWIGMGGGEPVK KEVIEQAVRRIMVGQEAEEMRNRAKELSQMAKRAVEEGGSS HNDFNSLIEDLRSRAH
GuUGT73P12_SCAFFOLD00629 (LC314779)	**Glycyrrhetinic acid 3-O-monoglucuronide glucuronosyltransferase**	**84**	ATGGACTCCTTTGGGGTTGAAGGTGATCACCAAGCCGACAC CACAGTGCTGAAGGCGGTTTTTCTTCCCTTCATCTCAAAAA GTCATCTCATCCGTGAGGTGGACAAAGCAAGGATCTTCGCC ATGCACGGCGTGGATGTCACCATCATCACCACGCCGGCCA ACGCTGCCACTTTCCAAACCTCCATTGACCGCGACTCCAGC CGCGGCCGCTCCATCAGAACGCACATCGTTCCGTTCCCCCA AGTCCCCGGTCTACCACAGGGACTCGAGAGACTCGACGCC GACACTCCTCAACACTTGCTCTCCAAGATCTACCATGGACT ATCCATTCTGCAAGAGCAGTTCCAACAACTGTTCCGTGAAA TGAGGCCAGATTTCATAGTCACTGACATGTACTACCCTTGG AGCGTCGATGCCGCCGCCGAGTTGGGGATTCCGAGGTTGGT TTGTAACGGTGGAAGCTACTTCGCTCAGTCAGCTGTTAACT CCGTTGAGCTATTTTCACCACAAGCCAAGGTTGATTCAAAT ACCGAGACTTTTCTGCTTCCTGGGTTACCCCATGAGGTTGA GATGACACGTTTGCAACTACCGGATTGGCTTAGAGGAGCA CCGAATGAGTACACCTATTTGATGAAGATGATCAAGGATTC AGAGAGGAAGAGTTATGGGTCATTGTTCAATAGCTTTTATG AGCTTGAAGGGACTTATGAGGAACATTACAAGAAAGCCAT GGGAACCAAGAGTTGGAGTGTGGGGCCAGTTTCTTTGTGG GTGAACCAAGATGCTTCTGATAAGGCTTGTAGGGGGGATG TTAAAGAAGGAAAAGGAGATGGGGTGGTGCTTACTTGGCT GGATTCTAAAACAGAGGACTCTGTTTTGTATGTGAGTTTTG GGAGCATGAACAAGTTCCCTAAAACTCAGCTTGTTGAGATA GCTCATGCCCTCGAAGATTCTGGCCATGATTTCATTTGGGT CGTTGGCAAAATTGAAGAAGGTGAAGGTGGTGCTGATTTTT TGAGGGAATTTGAGAAGAAAGTGAAAGAAAAAAACAGAG

			GTTATCTGATATGGGGTTGGGCACCACAGCTTCTGATTCTG GAGCATCCTGCGGTTGGAGCAGTGGTGACTCATTGTGGGTG GAACACCGTTATGGAAAGTGTGAATGCAAGTTTGCCATTGG CAACTTGGCCATTGTTTGCGGAGCAGTTCTTCAATGAGAAG CTAGTGGTTGATGTGGTGAAGATTGGTGTGCCAGTTGGGGT TAAGGAATGGAGAAATTGGAATGAGTTTGGGGATGAGGTT GTGAAGAGGGAGGACATAGGAAAGGCCATTGCTTTTTTGA TGGGTGGTGGGGATGAATCCTTGGAAATGAGGAAGAGGGT CAAGGTGCTCAGTGGTGCTACAAAGAAAGCTATTCAGGTT GGTGGGTCTTCTCACACCAAGTTGAAAGAACTCATAGAAG AGCTCAAGTCAATCAAGCTACAAAAGGTCAACAACAAATT AATGGAGGCAGTGGCTTAA
GuUGT73P12_SCAFFOLD00629 (LC314779)	**Glycyrrhetinic acid 3-O-monoglucuronide glucuronosyltransferase**	**85**	MDSFGVEGDHQADTTVLKAVFLPFISKSHLIREVDKARI FAMHGVDVTIITTPANAATFQTSIDRDSSRGRSIRTHIVPF PQVPGLPQGLERLDADTPQHLLSKIYHGLSILQEQFQQLF REMRPDFIVTDMYYPWSVDAAAELGIPRLVCNGGSYFA QSAVNSVELFSPQAKVDSNTETFLLPGLPHEVEMTRLQL PDWLRGAPNEYTYLMKMIKDSERKSYGSLFNSFYELEG TYEEHYKKAMGTKSWSVGPVSLWVNQDASDKACRGD VKEGKGDGVVLTWLDSKTEDSVLYVSFGSMNKFPKTQ LVEIAHALEDSGHDFIWVVGKIEEGEGGADFLREFEKKV KEKNRGYLIWGWAPQLLILEHPAVGAVVTHCGWNTVM ESVNASLPLATWPLFAEQFFNEKLVVDVVKIGVPVGVK EWRNWNEFGDEVVKREDIGKAIAFLMGGGDESLEMRK RVKVLSGATKKAIQVGGSSHTKLKELIEELKSIKLQKVN NKLMEAVA
GuCSLG Protein from Chinese liquorice _(Glyur003152s00037491.1_Glycyrrhiza uralensis) (SOAP5 homolog; Cellulose Synthase Like G)	Cellulose Synthase Like G (Glucuronic acid transferase)	***102***	MASFTLHTETVQSWLLLSRLHILLHLAVVLLLLYYRITR FPFHAPTLPWTLMTVGEAIMAVLWFFNQAFRWRPVSRS VMTEKLPSDAKLPGLDIFVCTLDPEKEPTVEVMNTLVSA LAMDYPPDKLSVYLSDDGAAPVTLYGVREASEFARVW VPFCKKYGIKSRCPKVFFSPSAEDEHLLRTDEFRSERDLI KAKYEKMQKNIEKFGSDAKNCRMVTDRPPRIEILIDQPD MPRVVYVSRERRPSLPHKFKGGALNTLLRVSGLISNGPY VLVVDCDMYCNDPSSAKQAMCFFLDPETSKSIAFVQFP QMFHNLGKKDIYDNQSRTAFKTMWQGMDGLRGPGLS GSGNYLNRSALLFGSPNQKDDYLDDAQNYLGKSTMYIE SLKAIRGQKTMKKNISRDEILREAQVLASCSYETNTEWG AEVGFSYGILLESSITGYLFHCRGWKSAYLYPKTPCFLG CAPTDIKEGMLQLVKWLSEYCLLGFSKYSPFTYGFSRMP IMPTLVYCFLTTTTLYSIVFILYGIVPQVCFLKGIPVFPKV TDPWFAVFATLYISTQIQHLIEVLSGDGSVAMWWDEQG IWILKSVTSVFAIIEAAKKGLGLNKKKFMLSNKAIDKEK LKKYEQGRFDFQGAALFMSPMVVLLIVNVVSFIGGIWR LFNAKDIEDMFGQLFLVSYVMALSYPIFEGIITMKSKSG
GuCSLG Gene from Chinese liquorice _(Glyur003152s00037491.1_Glycyrrhiza uralensis) (SOAP5 homolog;	Cellulose Synthase Like G (Glucuronic acid transferase)	**103**	ATGGCAAGCTTCACCCTTCACACAGAAACCGTTCAGT CATGGCTACTCCTCAGCAGACTTCACATACTGCTGCA CCTCGCAGTTGTACTGCTCCTCTTATACTACCGCATCA CACGTTTCCCCTTCCATGCTCCGACTCTACCGTGGACT CTGATGACCGTAGGTGAGGCTATTATGGCAGTGCTGT GGTTCTTCAACCAGGCCTTCCGGTGGCGGCCGGTGAG CCGCTCGGTGATGACGGAGAAGCTGCCCAGCGACGC GAAGCTGCCGGGGCTTGACATATTCGTGTGCACGCTT GACCCCGAGAAGGAGCCCACCGTGGAGGTGATGAAC ACTCTGGTCTCTGCCCTTGCCATGGACTACCCCCCTGA CAAGCTCTCCGTTTACCTCTCCGACGATGGCGCCGCC

Cellulose Synthase Like G)			CCGGTCACTCTTTACGGCGTGAGAGAGGCTTCTGAGT TCGCGAGGGTGTGGGTCCCTTTCTGCAAAAAGTATGG GATCAAGTCAAGGTGTCCCAAGGTTTTCTTCTCTCCCA GTGCTGAGGATGAACACCTTCTTCGCACCGACGAGTT CAGGTCAGAGCGAGACCTCATCAAGGCTAAATACGA GAAAATGCAGAAAAATATTGAGAAATTTGGTTCGGAT GCCAAAAATTGTCGTATGGTGACTGACAGACCTCCTC GGATCGAGATATTGATTGACCAACCAGACATGCCACG TGTTGTTTACGTGTCTCGGGAAAGAAGGCCATCACTC CCTCACAAGTTCAAAGGAGGAGCCCTCAATACATTGC TCAGAGTCTCAGGTCTAATCAGCAATGGGCCTTATGT ACTTGTAGTGGACTGTGATATGTATTGCAATGACCCA TCCTCAGCCAAACAAGCCATGTGTTTCTTTCTTGATCC TGAAACCTCTAAATCTATTGCATTTGTCCAATTCCCTC AAATGTTTCACAACCTTGGCAAAAAAGACATCTATGA CAATCAATCTAGGACTGCTTTTAAGACAATGTGGCAA GGGATGGATGGACTAAGAGGTCCTGGTCTTTCTGGCA GCGGTAATTACTTGAATAGAAGTGCATTACTATTTGG AAGTCCAAATCAAAAAGATGACTATCTGGATGATGCC CAAAACTACTTAGGCAAGTCTACCATGTACATAGAAT CACTAAAGGCCATTCGTGGACAAAAAACTATGAAAA AGAATATTTCAAGAGATGAAATTTTACGAGAAGCTCA AGTATTAGCCTCTTGTTCCTATGAGACAAACACAGAA TGGGGAGCAGAGGTAGGATTCTCATATGGCATCTTAC TGGAGAGTTCAATCACTGGCTATCTTTTCCACTGCAG AGGATGGAAATCAGCATATCTTTACCCAAAGACACCA TGTTTCTTAGGGTGTGCCCCAACTGACATCAAGGAAG GAATGCTCCAATTGGTGAAGTGGTTGTCTGAATACTG CTTGCTAGGATTCTCTAAATACAGCCCTTTCACTTATG GCTTTTCAAGAATGCCCATTATGCCTACCTTAGTCTAT TGCTTCTTGACAACAACAACCCTTTATTCCATTGTCTT CATCCTTTATGGCATTGTCCCCCAAGTTTGCTTCTTAA AAGGAATACCCGTGTTTCCAAAGGTCACAGACCCTTG GTTTGCAGTGTTTGCAACACTGTATATATCCACCCAG ATTCAACATTTGATAGAGGTCCTTTCTGGTGATGGCTC TGTGGCAATGTGGTGGGATGAACAGGGAATCTGGATT CTGAAGTCAGTCACTAGCGTGTTCGCAATCATAGAGG CAGCTAAGAAAGGGTTAGGATTGAACAAGAAGAAAT TCATGTTGTCAAACAAAGCAATTGACAAGGAGAAGCT CAAGAAGTATGAGCAAGGTAGGTTTGATTTCCAAGGT GCAGCTCTGTTCATGTCCCCAATGGTTGTGTTGCTCAT AGTGAACGTTGTTTCCTTCATTGGTGGCATATGGAGA CTATTCAATGCAAAGGATATTGAAGATATGTTTGGTC AGCTTTTCCTAGTTAGTTATGTAATGGCCCTTAGTTAT CCCATTTTTGAAGGGATAATAACCATGAAAAGCAAGA GTGGATAG

[001004] Expression of spinach *bAS* (*SoSOAP1*; SEQ ID NO: 45) , *GuCYP88D6* (SEQ ID NO: 76), *GuCYP72A154* (SEQ ID NO: 78), and *GuCSL* (SEQ ID NO: 80) in *N. benthamiana* resulted in formation of glycyrrhetinic acid 3-O-monoglucuronide (GA-3-GlcA) (**FIGURES 46A-46E**). On the other hand, previously reported *GuUGAT* (SEQ ID NO: 82) believed to catalyze the continuous two-step glucuronidation of glycyrrhetinic acid to yield glycyrrhizin failed to attach GlcA to the aglycone

in the assays performed here (G. Xu, W. Cai, W. Gao, C Liu, A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetinic acid to yield glycyrrhizin. New Phytol. 212, 123-35 (2016).).

[001005] Transient expression of GuUGT73P12 (SEQ ID NO: 84), which was previously shown to attach the second GlcA, together with *GuCSL* and other enzymes gave glycyrrhizin as a final product in *N. benthamiana* (**FIGURES 47A-47F**). The assays also showed that *GuCSL* (SEQ ID NO: 81) can perform addition of the second glucuronic acid moiety to GA-3-GlcA, but with very low efficiency.

[001006] Furthermore, it will be shown in a heterologous system (*N. benthamiana*) that the first glycosylation step in the biosynthesis of QS-21 from *Q. saponaria* may be performed by a CSLG which attaches glucuronic acid to the quillaic acid at position C-3.

[001007] *Summary*: The discovery of a new type of biosynthetic reaction performed by cellulose synthase like G enzymes provides a valuable strategy for engineering pathways of important specialized metabolites and their sustainable production in heterologous systems (plant or fungal).

Example 24: Structural Analysis of Cellulose Synthase Like G

[001008] *Objective*: To analyze the structure of Cellulose Synthase Like G (CSLG) and compare it with the classical Cellulose Synthase A (CESA).

[001009] *Methods*: See Materials and Methods above.

[001010] *Results*: Phylogenetic analysis showed that CSLGs are evolutionarily distinct from the classical Cellulose Synthase A (CESA) enzymes that mediate cellulose biosynthesis (approximately 30% amino acid sequence similarity). However, these two protein groups share several structural features. A predicted 3D model obtained by template-based tertiary structure modelling of the spinach SOAP5 CSLG protein showed that it consists of two domains traversing a lipid bilayer (TMD) at the amino-termini and an additional four TMDs at the carboxy terminal end (M. Källberg et al., Template-based protein structure modeling using the RaptorX web server. Nature Protocols. 7, 1511–1522 (2012).). A 445 amino acid long loop protruding from the lipid bilayer is located between the transmembrane domains (**FIGURE 48; FIGURE 49A**). The amino acid sequences of the spinach SOAP5 enzyme, the cellulose synthase like G enzyme from Arabidopsis, and the cellulose synthase A enzyme subunits from Arabidopsis and their orthologs in spinach are presented below in **Table 19**. In addition, other cellulose synthase like genes are presented from the A, B, and E families. The current understanding is that cellulose synthase like genes from A, B, and E families each have completely different activities. Until the work presented here, only few CSLs were known, and none from CSLG family, were characterized. Those that were analyzed are involved in sugar metabolism,

biosynthesis of beta-glucans and hemicelluloses. It was never shown that any CSL (despite family it belongs) can work on triterpenoids. Thus, the results presented throughout the examples provided surprising and unexpected activities for a cellulose like synthase.

[001011] **Table 12: Amino Acid and Nucleotide Sequences of *Arabidopsis* CSLG and CESA enzymes**

NAME	ENZYME ACTIVITY	SEQ ID NO:	AA SEQ
Arabidopsis CESA1	Cellulose Synthase A (subunit 1)	67	MEASAGLVAGSYRRNELVRIRHESDGGTKPLKNMN GQICQICGDDVGLAETGDVFVACNECAFPVCRPCYE YERKDGTQCCPQCKTRFRRHRGSPRVEGDEDEDDV DDIENEFNYAQGANKARHQRHGEEFSSSSRHESQPIP LLTHGHTVSGEIRTPDTQSVRTTSGPLGPSDRNAISSP YIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGW KLKQEKNMLQMTGKYHEGKGGEIEGTGSNGEELQM ADDTRLPMSRVVPIPSSRLTPYRVVIILRLIILCFFLQY RTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWY PINRETYLDRLAIRYDRDGEPSQLVPVDVFVSTVDPL KEPPLVTANTVLSILSVDYPVDKVACYVSDDGSAML TFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKID YLKDKIQPSFVKERRAMKREYEEFKVRINALVAKAQ KIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSG GLDTDGNELPRLIYVSREKRPGFQHHKKAGAMNALI RVSAVLTNGAYLLNVDCDHYFNNSKAIKEAMCFMM DPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDIN MKGLDGIQGPVYVGTGCCFNRQALYGYDPVLTEED LEPNIIVKSCCGSRKKGKSSKKYNYEKRRGINRSDSN APLFNMEDIDEGFEGYDDERSILMSQRSVEKRFGQSP VFIAATFMEQGGIPPTTNPATLLKEAIHVISCGYEDKT EWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWY GYHGRLRLLERIAYINTIVYPITSIPLIAYCILPAFCLIT DRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIED WWRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFT VTSKATDEDGDFAELYIFKWTALLIPPTTVLLVNLIGI VAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFL KGLLGRQNRTPTIVIVWSVLLASIFSLLWVRINPFVDA NPNANNFNGKGGVF
Arabidopsis CESA3	Cellulose Synthase A (subunit 3)	68	MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRF VACDICSFPVCRPCYEYERKDGNQSCPQCKTRYKRL KGSPAIPGDKDEDGLADEGTVEFNYPQKEKISERML GWHLTRGKGEEMGEPQYDKEVSHNHLPRLTSRQDT SGEFSAASPERLSVSSTIAGGKRLPYSSDVNQSPNRRI VDPVGLGNVAWKERVDGWKMKQEKNTGPVSTQAA SERGGVDIDASTDILADEALLNDEARQPLSRKVSIPSS RINPYRMVIMLRLVILCLFLHYRITNPVPNAFALWLV SVICEIWFALSWILDQFPKWFPVNRETYLDRLALRYD REGEPSQLAAVDIFVSTVDPLKEPPLVTANTVLSILAV DYPVDKVSCYVFDDGAAMLSFESLAETSEFARKWVP FCKKYSIEPRAPEWYFAAKIDYLKDKVQTSFVKDRR AMKREYEEFKIRINALVSKALKCPEEGWVMQDGTP WPGNNTGDHPGMIQVFLGQNGGLDAEGNELPRLVY VSREKRPGFQHHKKAGAMNALVRVSAVLTNGPFILN

			LDCDHYINNSKALREAMCFLMDPNLGKQVCYVQFP QRFDGIDKNDRYANRNTVFFDINLRGLDGIQGPVYV GTGCVFNRTALYGYEPPIKVKHKKPSLLSKLCGGSR KKNSKAKKESDKKKSGRHTDSTVPVFNLDDIEEGVE GAGFDDEKALLMSQMSLEKRFGQSAVFVASTLMEN GGVPPSATPENLLKEAIHVISCGYEDKSDWGMEIGWI YGSVTEDILTGFKMHARGWRSIYCMPKLPAFKGSAPI NLSDRLNQVLRWALGSVEILFSRHCPIWYGYNGRLK FLERFAYVNTTIYPITSIPLLMYCTLLAVCLFTNQFIIP QISNIASIWFLSLFLSIFATGILEMRWSGVGIDEWWRN EQFWVIGGVSAHLFAVFQGILKVLAGIDTNFTVTSKA SDEDGDFAELYLFKWTTLLIPPTTLLIVNLVGVVAGV SYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLM GRQNRTPTIVVVWSVLLASIFSLLWVRIDPFTSRVTGP DILECGINC
Arabidopsis CSLG1	Cellulose Synthase Like G	69	METHRKNSVVGNILHTCHPCRRTIPYRIYAIFHTCGII ALMYHHVHSLVTANNTLITCLLLLSDIVLAFMWATT TSLRLNPVHRTECPEKYAAKPEDFPKLDVFICTADPY KEPPMMVVNTALSVMAYEYPSDKISVYVSDDGGSSL TFFALIEAAKFSKQWLPFCKKNNVQDRSPEVYFSSES HSRSDEAENLKMMYEDMKSRVEHVVESGKVETAFI TCDQFRGVFDLWTDKFSRHDHPTIIQVLQNSETDMD NTRKYIMPNLIYVSREKSKVSPHHFKAGALNTLLRVS GVMTNSPIILTLDCDMYSNDPATLVRALCYLTDPEIK SGLGYVQFPQKFLGISKNDIYACENKRLFIINMVGFD GLMGPTHVGTGCFFNRRAFYGPPYMLILPEINELKPY RIADKSIKAQDVLSLAHNVAGCIYEYNTNWGSKIGFR YGSLVEDYYTGFMLHCEGWRSVFCNPKKAAFYGDS PKCLVDLVGQQIRWAVGLFEMSFSKYSPITYGIKSLD LLMGLGYCNSPFKPFWSIPLTVYGLLPQLALISGVSV FPKASDPWFWLYIILFFGAYAQDLSDFLLEGGTYRK WWNDQRMLMIKGLSSFFFGFIEFILKTLNLSTPKFNV TSKANDDDEQRKRYEQEIFDFGTSSSMFLPLTTVAIV NLLAFVWGLYGILFCGGELYLELMLVSFAVVNCLPI YGAMVLRKDDGKLSKRTCFLAGNLHVGSYCVKLLR PQVTSPLRLIHNNNTSGWFKRKKHNMNESV
Spinach CESA1	cellulose synthase A catalytic subunit 1 [UDP-forming]	70	MEATGGMVAGSYKRNELVRIRHDSTDSGSKSLKNL DGQICQICGDTVGVTSNGGVFVACNECAFPVCRPCY EYERKDGNQCCPQCKTRYKRQKGSLRVEGDDEEED VDDLDNEFNYERGTSKARHQWQGEDVDLSSSSRHG SQPIPLLTNGQVVSGEIPSATPDNQSVRSTSGPIGPEKR GNHSLPYIDPCLPVPVRIVDPSKDLNSYGLGSVDWKE RVESWKLKQEKNMTHTGNRYSEGKGGDVEGSGSNG EELQLADDVRQPMSRIVPIPSSHLTPYRAVIIFRLIILVF FLQFRITHPVEDAYPLWLTSVICEIWFAMSWILDQFP KWYPINRETYLDRLAFRHDREGEPSQLAPIDVFVSTV DPLKEPPIITANTVLSILAVDYPVDKVSCYVSDDGSA MLTFEGLSETAEFARKWVPFCKKFSIEPRAPEFYFQQ KIDYLKDKIQPSFVKERRAMKREYEEFKVRINALVA KAQKVPEEGWTMQDGTAWPGNNPRDHPGMIQVFL GHSGGLDMDGNELPRLVYVSREKRPGFQHHKKAGA MNALIRVSAVLTNGAYILNVDCDHYFNNSKCLKEA MCFMMDPALGKKVCYVQFPQRFDGIDLHDRYANRN IVFFDINMKGQDGIQGPVYVGTGCCFNRQALYGYDP VLTEEDFEPNFIIKNCFGSRKKGKSGNKKYMDKKRG PKRSESSIPIFNMEDIEEGVEGYEDEKSLLMSQKRLEK

			RFGQSPVFIAATFMEMGGIPPTTNPATLLKEAIHVISC GYEDKSEWGKEIGWIYGSVTEDILTGFKMHARGWM SIYCMPPRPAFKGSAPLNLSDRLNQVLRWALGSIEIM LSRHCPIWYGYKGRLRFLERLAYINTVVYPLTSIPLIA YCILPAICLLTNKFIIPTLSNFASILFIMLFMSIAATGILE LRWSGVSIEDWWRNEQFWVIGGTSAHLFAVFQGLL KVLAGIDTNFTVTSKAADEDGDFAELYIFKWTALLIP PTTVLIVNLVGVVAGVSYAINSGYQSWGPLFGKLFFS FWVIAHLYPFLKGLLGRQNRTPTIVIVWSVLLASIFSL LWVRINPFTTDAEKAAAGNQCGINC
Spinach CESA3	cellulose synthase A catalytic subunit 3 [UDP-forming]	71	MMEDSQSGVKPTKQANEQVCQICSDNIGTTVDGEPF VACDVCSFPVCRACYEYERKDGTQSCPQCKTRYKR QKGSPAIHGEKVEDSDVEDVVSDVNEPLGSSILKEKP QERMLGWHMNHGQSGELGPPTYDKEAPISHIPRLAT GRTVSGDLSAASPGRFSMPSPGASTGANIRVSREFAS PGFGNVAWKERIDGWKMKQEKSTGPPSVSHAPSEG RFANDIDASTEIAMDDPLLNDETRQPLSRKVPIPSSRI NPYRMVIVLRLAVLGIFLHYRVTNPVPNAYALWLIS VICEIWFAFSWILDQFPKWLPINRETYLDRLALRYDR EGEPSQLAAVDIFVSTVDPLKEPPLVTANTVLSILAVD YPVDKVSCYVSDDGAAMLTFEALSETSEFARKWVPF TKKYNIEPRAPEWYFSQKIDYLKDKVQTTFVKDRRA MKREYEEFKIRINGLVAKATKVPEEGWVMQDGTPW PGNNTRDHPGMIQVFLGQSGGLDTDGNELPRLVYVS REKRPGFTHHKKAGAMNSLVRVSAVLTNGPFMLNL DCDHYINNSKALREAMCFMMDPNLGKYCCYVQFPQ RFDGIDRNDRYANRNTVFFDINLRGLDGIQGPVYVG TGCVFNRTALYGYEPPIKPKPKKKGILSSCFGGSRKK SSKKDSKKKSKHADPTVPIFNLEDIEEGVEGTGFDDE KSLLMSQISLEKRFGKSEVFVASTLMENGGVPQSATP DTLLKEAIHVISCGYEDKTDWGAEIGWIYGSVTEDIL TGFKMHARGWRSIYCMPKLAAFKGSAPINLSDRLNQ VLRWALGSVEILFSRHCPLWYGYGGRLKWLERFAYI NTTIYPLTSIPLLAYCTLPAVCLLTGKFIIPQISNLASV WFLSLFLSIFATGILEMRWSGVGIDEWWRNEQFWVI GGISAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGD FTELYLFKWTTLLIPPTTILIVNLVAVVAGISYAINSGY QSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTPT IVVVWSILLASIFSLLWVRVDPFTTRVTGPDVHICGIN C
ATCSLA1_At4g16590		86	ATGTCTCTATTTCTGAAGCCCTTCCTCTTCCTATAC GACACCACTCTTAGTCTTCTCTTACTTCTGTTCAAT GGATGGAGTCTTGAGGATACAGCAGCAGCCCAAA AGAGGCGTGAAGCAGACAAAAATGCTGCAGAAAC TGAATGGATCCAACTCCAATACTTGTGGACCAAAA CAAGGAGTGTTGTACTACTTCCCGTTTTCAAGGGT TTGGTGGTTATGTGTTTGGTTCTATCCATTATAGTG TTCTTCGAGAGTTTTTACATGAACTTTGTGATACTC TTCGTCAAGTTATTTAAACGTAAACCCCATAAAGT GTACAAATGGGAGGCCATGCAAGAAGATGTTGAG GTTGGACCCGATAACTACCCAATGGTTCTTATCCA AATACCAATGTACAATGAAAAAGAGGAAGGTGTG GACGTAGAGATTGCAAAATGGCAAAGCCAAGGCA TAAACATAAGGTGTGAAAGGAGAGATAACAGGAA CGGCTACAAAGCCGGAGCTATGAAAGAAGCTCTT ACGCAGAGCTACGTCAAGCAATGCGACTTCGTAGC

			AGTCTTCGATGCTGATTTCCAACCCGAGCCCGATT ATCTCATCCGCGCTGTCCCTTTCCTTGTCCACAACC CTGACGTTGCTCTAGTTCAAGCCCGATGGATATTT GTTAACGCGAACAAATGCTTGATGACGAGGATGC AAGAGATGTCTCTCAACTATCATTTCAAAGTGGAA CAAGAATCAGGGTCGACTAGACATGCTTTCTTCGG GTTTAATGGAACCGCGGGTGTATGGAGAATATCGG CAATGGAAGCAGCAGGAGGATGGAAATCAAGGAC CACAGTAGAGGACATGGACTTGGCTGTTCGTGTTG GTCTTCATGGCTGGAAATTTGTCTACCTTAACGAC CTCACGGTGAGAAACGAGCTTCCAAGCAAATTTAA GGCCTACAGATTCCAGCAACATAGGTGGTCCTGTG GACCGGCGAATCTATTTAGAAAAATGACGATGGA GATCATTTTCAATAAGAGAGTATCAATTTGGAAGA AGTTTTATGTGATCTACAGCTTTTTCTTCGTAAGGA AAGTGGCGGTACACTTCTTGACATTCTTCTTCTACT GTATAATTGTGCCAACAAGTGTCTTCTTCCCTGAA ATCCACATCCCATCTTGGTCTACCATTTACGTTCCC TCTTTGATCAGTATCTTCCACACCCTGGCAACTCCA AGATCCTTCTACCTCGTGATATTTTGGGTCTTGTTC GAGAATGTAATGGCTATGCATCGAACCAAAGGTA CGTGCATTGGCCTACTTGAAGGAGGAAGAGTAAA CGAATGGGTTGTGACCGAAAAACTAGGAGATGCTT TGAAGAGTAAGCTACTCTCTCGGGTAGTCCAAAGA AAATCTTGTTATCAAAGAGTGAATTCCAAGGAAGT GATGGTGGGGGTATACATATTAGGATGTGCACTCT ATGGCCTGATCTATGGGCACACATGGTTACATTTC TATCTTTTTCTTCAGGCCACAGCCTTTTTCGTCTCC GGTTTTGGTTTTGTCGGAACCTAA
ATCSLA1_At4g16590		87	MSLFLKPFLFLYDTTLSLLLLLFNGWSLEDTAAAQKR READKNAAETEWIQLQYLWTKTRSVVLLPVFKGLV VMCLVLSIIVFFESFYMNFVILFVKLFKRKPHKVYKW EAMQEDVEVGPDNYPMVLIQIPMYNEKEEGVDVEIA KWQSQGINIRCERRDNRNGYKAGAMKEALTQSYVK QCDFVAVFDADFQPEPDYLIRAVPFLVHNPDVALVQ ARWIFVNANKCLMTRMQEMSLNYHFKVEQESGSTR HAFFGFNGTAGVWRISAMEAAGGWKSRTTVEDMDL AVRVGLHGWKFVYLNDLTVRNELPSKFKAYRFQQH RWSCGPANLFRKMTMEIIFNKRVSIWKKFYVIYSFFF VRKVAVHFLTFFFYCIIVPTSVFFPEIHIPSWSTIYVPSL ISIFHTLATPRSFYLVIFWVLFENVMAMHRTKGTCIGL LEGGRVNEWVVTEKLGDALKSKLLSRVVQRKSCYQ RVNSKEVMVGVYILGCALYGLIYGHTWLHFYLFLQA TAFFVSGFGFVGT
ATCSLB1_AT2G32610		88	ATGGCGGATTCAAGCTTTTCTCTTCCTCCTCTTTGT GAAAGGATCTCATACACGAACTATTTTCTAAGAGC TGTATATCTCACGGTTCTAGGCCTTTTCTTTTCTCTT CTCTTGCACCGAATCCGACATACGAGCGAATACGA CAACGTTTGGCTCGTGGCTTTCTTTTGTGAATCTTG TTTCTTCTTGGTATGTCTGCTTATTACTTGCCTAAA ATGGAGTCCTGCTGATACTAAACCCTTTCCTGATA GACTTGATGAAAGGGTTCATGACCTTCCTTCGGTG GATATGTTCGTGCCCACAGCAGATCCGGTTCGAGA GCCACCGATTATGGTTGTGGACACCGTGCTTTCGC TGTTAGCTGTAAATTATCCGGCAAATAAACTAGCT TGTTATGTGTCGGACGATGGATGCTCACCTCTCAC

			TTATTTCTCTCTCAAGGAAGCTTCTAAGTTCGCCAA GATTTGGGTACCGTTCTGCAAAAAGTACAACACTA GAGTTAGAGCTCCTTCTAGATATTTTCTGAAACCT ATAAGCGTCGCAACAGAGGATTATGAATTCAATAG AGACTGGGAAAAGACGAAGAGGGAGTACGAGAA GTTGAGGCGGAAAGTGGAAGATGCCACCGGAGAT TCTCATATGTTGGATGTAGAAGATGATTTTGAAGC ATTCTCAAACACAAAACCAAATGATCATTCAACTC TAGTTAAGGTGGTATGGGAGAACAAGGGAGGTGT AGGAGACGAGAAAGAGATCCCTCATATCATATAC ATATCAAGAGAGAAAAGACCAAATTATGTTCATA ATCAAAAATGTGGAGCCATGAACTTTCTGGCAAGA GTGTCAGGGTTGATGACAAACGCACCATACATCTT GAACGTGGATTGCGACATGTATGCCAATGATGCAG ATGTAGTCCGACAAGCAATGTGTATACTTCTGCAA GAATCATTAAATATGAAACATTGTGCTTTTGTTCA ATTCCGTCAAGAATTCTATGATTCAAGCACCGAGC TAATAGTCGTCCTACAATCACATTTGGGACGAGGA ATCGCGGGAATCCAAGGACCGATATATATAGGATC AGGATGCGTCCACACGAGAAGAGTTATGTATGGTT TATCTCCAGACGATTTCGAAGTTGATGGAAGTCTT TCTTCAGTTGCTACAAGGGAGTTTTTGGTTAAGGA TAGTTTAGCGAGAAGATTTGGTAATTCTAAAGAGA TGATGAAATCAGTGGTTGATGCAATACAAAGAAAT CCAAATCCACAAAATATACTTACAAACTCCATAGA AGCGGCTCGAGAAGTGGGACATTGTCAGTACGAG TACCAAACCAGCTGGGGAAACACCATCGGCTGGTT ATATGATTCAGTGGCGGAAGATTTAAACACGAGTA TCGGAATACATTCGAGAGGTTGGACTAGCTCATAC ATTTCTCCGGATACACCTGCATTTCTTGGATCTATG CCGGCAGGAGTACCCGAGGCGTTACTCCAGCAGC GTCGATGGGCGACAGGATGGATCGAAATCCTTTTC AACAAGCAAAGTCCGTTGCGAGGATTGTTTAGCAA GAAAATAAGATTCCGACAACGATTAGCTTATCTTT GCATTATCACCTGTCTAAGGTCAATCCCTGAGCTT ATTTATTGTCTCCTTCCTGCTTATTGCCTACTCCAC AACTCTACCTTATTCCCCAAGGGACTTTATTTAGGC ATAACTGTCACACTTGTTGGGATACATTGTCTCTAT ACTCTATGGGAATTTATGAGCCTTGGTTATTCCGTA CAATCGTGGCTAGTCTCCCAATCAGTTTGGAGAAT AGTAGCCACTAGTAGTTGGTTATTTAGCATCTTTG ATATCACACTCAAGCTTCTTGGCATCTCGGAAACG GTGTTCATAATCACTAAAAAGACTGTGGCTGGGAC CAAGTCAGCATTAGGGTCTGGACCCTCTCAAGGAG AAGACGTTGGTCCAAACTCAGACTTGTTTAAATTT GAATTTGATGGCTCACTTTGTTTCTTGCCTGGCACA TTTATTGTGTTGGTGAATATAGCCGCTCTAGCTGTT TTTTCTGTGGGTCTACAACGGTCGAGTTACAGCCA TGAAGGAGGTGGTTCGGGTCTGGCAGAGGCTTGCG GATGTGTTTTGGTAATGATGTTGTTCCTTCCATTTC TAATGGGTTTGTTTAAGAAAGGAAAATATGGAACC CCATTGTCTACTCTCTCTATAGCTGGCTTTTTAGCA GTTTTATTTGTTGTTTTCTCTGTTTGA
ATCSLB1 _AT2G326 10		89	MADSSFSLPPLCERISYTNYFLRAVYLTVLGLFFSLLL HRIRHTSEYDNVWLVAFFCESCFFLVCLLITCLKWSP ADTKPFPDRLDERVHDLPSVDMFVPTADPVREPPIM

			VVDTVLSLLAVNYPANKLACYVSDDGCSPLTYFSLK EASKFAKIWVPFCKKYNTRVRAPSRYFLKPISVATED YEFNRDWEKTKREYEKLRRKVEDATGDSHMLDVED DFEAFSNTKPNDHSTLVKVVWENKGGVGDEKEIPHII YISREKRPNYVHNQKCGAMNFLARVSGLMTNAPYIL NVDCDMYANDADVVRQAMCILLQESLNMKHCAFV QFRQEFYDSSTELIVVLQSHLGRGIAGIQGPIYIGSGC VHTRRVMYGLSPDDFEVDGSLSSVATREFLVKDSLA RRFGNSKEMMKSVVDAIQRNPNPQNILTNSIEAAREV GHCQYEYQTSWGNTIGWLYDSVAEDLNTSIGIHSRG WTSSYISPDTPAFLGSMPAGVPEALLQQRRWATGWI EILFNKQSPLRGLFSKKIRFRQRLAYLCIITCLRSIPELI YCLLPAYCLLHNSTLFPKGLYLGITVTLVGIHCLYTL WEFMSLGYSVQSWLVSQSVWRIVATSSWLFSIFDITL KLLGISETVFIITKKTVAGTKSALGSGPSQGEDVGPNS DLFKFEFDGSLCFLPGTFIVLVNIAALAVFSVGLQRSS YSHEGGGSGLAEACGCVLVMMLFLPFLMGLFKKGK YGTPLSTLSIAGFLAVLFVVFSV
ATCSLE1_AT1G55850		90	ATGGTAAACAAAGACGACCGGATTAGACCGGTTC ATGAAGCCGACGGTGAACCGCTTTTTGAGACTAGG AGAAGAACCGGTAGAGTGATTGCGTACCGGTTTTT CTCAGCCTCGGTTTTCGTGTGTATCTGTTTGATTTG GTTCTACAGAATTGGTGAGATTGGTGATAACCGTA CCGTTTTAGATCGATTAATCTGGTTTGTTATGTTTA TTGTGGAGATTTGGTTCGGTTTATATTGGGTAGTCA CACAATCTTCCCGGTGGAATCCGGTTTGGCGATTT CCCTTCTCCGATAGACTCTCTCGGAGATACGGAAG CGACCTTCCGAGGCTCGACGTCTTCGTTTGCACGG CGGATCCGGTGATTGAGCCGCCGTTGTTGGTGGTA AACACAGTCTTATCTGTGACGGCTCTTGACTACCC ACCGGAGAAACTCGCCGTTTATCTCTCAGATGACG GTGGTTCTGAGCTGACGTTCTATGCTCTCACGGAG GCAGCTGAGTTTGCTAAAACTTGGGTTCCCTTCTG CAAGAAGTTCAACGTTGAGCCAACATCTCCCGCTG CTTACTTGTCTTCCAAGGCAAACTGTCTTGATTCTG CGGCTGAGGAGGTGGCTAAGCTGTATAGAGAAAT GGCGGCGAGGATTGAAACGGCGGCGAGACTGGGA CGAATACCGGAGGAGGCGCGGGTGAAGTACGGTG ACGGGTTTTCACAGTGGGATGCTGACGCTACTCGA AGAAACCATGGAACCATTCTTCAAGTTTTGGTAGA TGGAAGAGAAGGGAATACAATAGCAATACCAACG TTGGTGTATTTATCAAGAGAAAAGAGACCTCAACA TCATCATAACTTCAAGGCTGGAGCAATGAACGCAT TGCTGAGGGTTTCTTCGAAAATTACTTGTGGGAAA ATCATACTAAACTTGGACTGTGATATGTACGCAAA CAACTCAAAGTCAACACGCGACGCGCTCTGCATCC TCCTCGATGAGAAAGAGGGAAAAGAGATTGCTTTC GTGCAGTTTCCGCAGTGTTTTGACAATGTTACAAG AAATGATTTGTATGGAAGCATGATGCGAGTAGGA ATTGATGTGGAATTTCTTGGATTGGATGGAAATGG TGGTCCGTTATACATTGGAACTGGATGCTTTCACA GAAGAGATGTGATCTGTGGAAGAAAGTATGGAGA GGAAGAAGAAGAAGAAGAATCTGAGAGAATTCAC GAAAATTTAGAGCCTGAGATGATTAAGGCTCTCGC GAGCTGCACTTATGAGGAAAACACTCAATGGGGA AAGGAGATGGGTGTGAAATATGGTTGCCCGGTAG

			AGGATGTAATAACTGGTTTGACGATTCAGTGTCGC GGATGGAAATCAGCCTACCTGAACCCGGAAAAGC AAGCATTTCTCGGGGTAGCGCCGACCAATTTGCAT CAAATGCTAGTGCAGCAGAGGAGATGGTCAGAGG GAGACTTTCAGATTATGCTTTCGAAGTATAGTCCG GTTTGGTATGGAAAAGGAAAGATCAGTTTAGGACT GATACTTGGTTACTGTTGCTATTGTCTTTGGGCTCC ATCTTCACTACCTGTGCTCATTTACTCTGTTTTGAC TTCTCTCTGTCTCTTCAAAGGCATTCCTCTGTTTCC AAAGGTCTCGAGCTCGTGGTTTATTCCGTTTGGAT ACGTCACTGTTGCAGCTACCGCATATAGCCTAGCC GAGTTCTTGTGGTGCGGAGGGACGTTCCGTGGATG GTGGAACGAGCAAAGGATGTGGCTTTATAGAAGA ACAAGCTCGTTTCTTTTCGGATTTATGGACACGATT AAGAAGCTACTTGGAGTTTCTGAGTCTGCGTTTGT GATCACAGCAAAAGTAGCAGAAGAAGAAGCAGCA GAGAGATACAAGGAAGAGGTAATGGAGTTTGGAG TGGAGTCTCCCATGTTTCTCGTCCTCGGAACACTCG GTATGCTCAATCTCTTCTGCTTCGCCGCAGCGGTTG CGAGACTTGTTTCCGGAGACGGTGGAGATTTGAAA ACAATGGGGATGCAATTTGTGATAACAGGAGTACT AGTTGTCATAAACTGGCCTCTGTATAAAGGTATGT TGTTGAGGCAAGACAAAGGAAAGATGCCAATGAG CGTTACAGTTAAATCAGTTGTTTTAGCTTTATCTGC CTGTACCTGTTTAGCGTTTTTGTAA
ATCSLE1_AT1G55850		91	MVNKDDRIRPVHEADGEPLFETRRRTGRVIAYRFFSA SVFVCICLIWFYRIGEIGDNRTVLDRLIWFVMFIVEIW FGLYWVVTQSSRWNPVWRFPFSDRLSRRYGSDLPRL DVFVCTADPVIEPPLLVVNTVLSVTALDYPPEKLAVY LSDDGGSELTFYALTEAAEFAKTWVPFCKKFNVEPT SPAAYLSSKANCLDSAAEEVAKLYREMAARIETAAR LGRIPEEARVKYGDGFSQWDADATRRNHGTILQVLV DGREGNTIAIPTLVYLSREKRPQHHHNFKAGAMNAL LRVSSKITCGKIILNLDCDMYANNSKSTRDALCILLDE KEGKEIAFVQFPQCFDNVTRNDLYGSMMRVGIDVEF LGLDGNGGPLYIGTGCFHRRDVICGRKYGEEEEEEES ERIHENLEPEMIKALASCTYEENTQWGKEMGVKYGC PVEDVITGLTIQCRGWKSAYLNPEKQAFLGVAPTNL HQMLVQQRRWSEGDFQIMLSKYSPVWYGKGKISLG LILGYCCYCLWAPSSLPVLIYSVLTSLCLFKGIPLFPK VSSSWFIPFGYVTVAATAYSLAEFLWCGGTFRGWW NEQRMWLYRRTSSFLFGFMDTIKKLLGVSESAFVITA KVAEEEAAERYKEEVMEFGVESPMFLVLGTLGMLN LFCFAAAVARLVSGDGGDLKTMGMQFVITGVLVVIN WPLYKGMLLRQDKGKMPMSVTVKSVVLALSACTCL AFL
ATCSLG1_AT4G24010		92	ATGGAGACTCATAGAAAGAACTCGGTCGTCGGCA ACATCCTCCACACGTGTCATCCTTGCCGGCGCACC ATTCCATATAGAATCTACGCCATATTTCACACGTG TGGCATCATAGCTCTCATGTATCACCATGTACACT CACTTGTCACAGCAAACAACACTCTTATAACATGT CTTCTTCTCCTCTCCGATATTGTTCTCGCCTTCATGT GGGCAACCACAACTTCCCTCCGCTTAAACCCGGTT CATCGGACCGAGTGCCCGGAGAAATATGCAGCTA AACCGGAGGACTTTCCAAAGCTGGACGTGTTTATA TGCACGGCTGATCCGTACAAGGAGCCTCCAATGAT

			GGTTGTTAACACCGCTTTATCGGTGATGGCTTACG AGTATCCGTCAGATAAGATCTCGGTGTATGTATCG GACGATGGAGGATCGTCGTTGACTTTCTTTGCTCTT ATTGAAGCTGCTAAGTTCTCTAAGCAGTGGTTGCC CTTTTGCAAGAAGAATAATGTTCAAGATCGGTCTC CTGAAGTTTATTTCTCTTCAGAGTCACATTCTCGAA GTGATGAAGCTGAAAACCTTAAGATGATGTACGA AGACATGAAGAGTAGAGTAGAACATGTGGTGGAG AGTGGAAAAGTTGAAACTGCGTTTATCACATGCGA CCAATTTCGTGGGGTATTCGATTTGTGGACCGACA AATTCAGTCGTCATGACCATCCCACAATTATTCAG GTGTTGCAAAATAGCGAGACAGATATGGACAATA CCAGAAAATATATAATGCCAAACCTAATCTATGTT TCAAGAGAGAAGAGTAAAGTTTCACCACATCATTT CAAAGCTGGTGCTCTTAATACTTTGCTACGAGTAT CAGGGGTGATGACAAATTCACCGATCATTCTAACA CTAGACTGTGATATGTATTCGAACGACCCGGCAAC ACTGGTTCGTGCTTTGTGCTATTTAACAGATCCTGA AATCAAATCCGGTTTAGGATATGTGCAGTTTCCTC AGAAATTTCTAGGAATAAGCAAAAATGATATATAT GCTTGTGAAAACAAACGCCTCTTCATTATTAATAT GGTTGGGTTTGATGGTCTAATGGGTCCAACTCATG TGGGAACTGGTTGTTTCTTTAATCGACGAGCTTTCT ATGGACCTCCATATATGTTGATTTTACCGGAGATA AATGAACTAAAGCCTTATCGGATTGCGGATAAGTC TATCAAAGCCCAAGATGTTTTGTCATTAGCACACA ATGTAGCAGGATGTATCTATGAGTACAATACCAAT TGGGGATCCAAGATTGGATTCAGATATGGGTCATT AGTAGAAGACTACTACACAGGGTTTATGCTCCATT GTGAAGGATGGAGATCAGTATTTTGCAACCCAAAA AAAGCTGCATTTTATGGAGATTCCCCAAAGTGCCT AGTTGATCTTGTGGGTCAACAAATCCGTTGGGCAG TTGGGCTTTTCGAAATGTCCTTTTCAAAGTATAGCC CAATTACCTATGGAATCAAGTCACTGGACCTTTTA ATGGGTTTAGGTTATTGCAACTCTCCGTTTAAGCC ATTTTGGTCAATTCCTCTGACCGTCTATGGACTTTT ACCACAGCTTGCACTCATTTCTGGAGTTAGTGTCTT CCCCAAGGCATCTGATCCGTGGTTTTGGCTTTACAT CATTTTATTCTTTGGGGCTTATGCCCAAGATCTATC AGACTTTTTATTGGAAGGAGGAACTTATCGGAAAT GGTGGAACGATCAAAGAATGTTGATGATAAAAGG ACTCTCTTCATTCTTCTTTGGTTTTATAGAGTTCATT CTCAAAACCCTAAACCTCTCCACACCTAAGTTCAA CGTCACCAGTAAAGCCAATGATGATGACGAACAG AGGAAGCGGTACGAGCAAGAAATCTTTGATTTCGG AACCTCTTCGTCCATGTTCTTGCCCTTGACCACGGT TGCCATAGTGAATCTGCTTGCTTTTGTCTGGGGGCT TTATGGTATTCTCTTCTGCGGAGGAGAACTCTACCT TGAGCTGATGCTGGTGAGCTTCGCGGTGGTGAATT GCTTACCGATCTACGGGGCTATGGTGTTGAGGAAA GATGATGGAAAATTATCAAAAAGAACTTGTTTCTT AGCTGGGAACCTCCACGTTGGTTCTTATTGTGTCA AGTTACTTCGTCCTCAAGTAACTTCACCCCTTAGGT TAATTCACAACAATAATACGTCTGGCTGGTTCAAG CGGAAGAAACACAATATGAATGAATCTGTGTAA

[001012] Three channels were also detected that could possibly form the passage of the triterpenoid aglycone into the enzyme active site and a channel that could be the exit path for the glucuronidated product (**FIGURES 49A-49C**). In cellulose synthase A (CESA) proteins, transmembrane domains (TMDs) form a pore that allows passing of polysaccharide chains through the membrane to the extracellular space. The out-of-membrane domain in CESAs holds the enzyme active site, as well as four motifs that are conserved among all such proteins and predicted to be involved in substrate and/or acceptor binding, i.e., DD, DCD, ED and QVLRW (L. Sethaphong et al., Tertiary model of a plant cellulose synthase, Proc. Natl. Acad. Sci. U.S.A. 18, 7512-7 (2013).) (**FIGURES 48**- red boxes). Alignment of amino acid sequences of CESA and SOAP5 from spinach showed that ED (in SOAP5: ES) and QVLRW (in SOAP5: QLIKW) motifs are not conserved in SOAP5 (**FIGURE 48**).

[001013] To check if the presence of CESA type amino acid motifs can affect SOAP5 activity, three mutated version of SOAP5 were created, as follows: (i) serine 442 substituted by aspartic acid (S442D), (ii) lysine 483 swapped with arginine (K483R) and (iii) both substitutions (S442D/K483R). Expression of mutated proteins together with SOAP1-4 in *N. benthamiana* demonstrated that modification of these motifs in SOAP5 alters enzyme performance. Wild type SOAP5 converts the entire medicagenic acid pool to MA-3-GlcA (MA-3-GlcA : MA: ratio 0 : 1), SOAP5-S442D or SOAP5-K483R could still effectively glucuronidate MA (MA-3-GlcA : MA ratio 0.03 : 1 and 0.09 : 1, respectively), while SOAP5-S442D/K483R was only able to partially process MA (MA-3-GlcA : MA ratio 0.90 : 1) (**FIGURE 50A and 50B**). Our observations indicate that change of the neutral amino acid to a negatively charged one causes steric hindrance in the enzyme active site in addition to affecting binding of the anionic substrate (UDP-GlcA), thus simultaneously reducing the enzymes efficiency. Changes of the amino acids (S442D and K483R) in the active site during the enzymes evolution could have been crucial for accepting UDP-GlcA instead of UDP-Glc and resulted in neofunctionalization of CESA proteins.

[001014] *Summary*: It appears that differences in key amino acid residues of CESA result in neofunctionalization of the enzyme to a CSLG enzyme activity. Another feature of CSLG that is crucial for its new triterpenoid related function was localization to endoplasmic reticulum (ER).

Example 25: Subcellular Localization of the Triterpenoid Saponin Biosynthetic Pathway

[001015] *Objective*: To investigate the subcellular localization of the triterpenoid saponin biosynthetic pathway.

[001016] *Methods*: See Materials and Methods above.

[001017] *Results*: The subcellular localization of the triterpenoid saponin biosynthetic pathway was

studies. The ER membrane system is a pertinent environment for the production of specialized metabolites such as steroids and triterpenoids. Many enzymes involved in triterpenoid biosynthesis, including squalene synthase, squalene epoxidase, oxidosqualene cyclases and most of the known CYP450s are membrane bound and operate within the ER compartment (C.A. Hasemann et al., Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure. 1, 41-62 (1995); K. B. Linscott, T. D. Niehaus, X. Zhuang, S. A. Bell, J. Chappell, Mapping a kingdom-specific functional domain of squalene synthase. BBA - Molecular and Cell Biology of Lipids. 1861, 1049–1057 (2016); M. Christen et al., Structural insights on cholesterol endosynthesis: Binding of squalene and 2,3-oxidosqualene to supernatant protein factor. Journal of Structural Biology. 190, 261–270 (2015); R. Thoma et al., Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase, Nature. 7013, 118-22 (2004).).

[001018] Transient expression of SOAP5 fused to a fluorescent reporter (SOAP5:mRFP) together with a cellular compartment marker (ER marker in fusion with GFP) demonstrated its localization to the ER network (**FIGURES 51A-51F**). Previously it was shown that the microsomal fraction from germinating soybean seeds contains a protein able to glucuronidate soyasapogenol B, supporting transmembrane localization of SOAP5 (Y. Kurosawa, H. Takahara, M Shiraiwa, UDP-glucuronic acid:soyasapogenol glucuronosyltransferase involved in saponin biosynthesis in germinating soybean seeds. Planta. 215, 620-9 (2002)). Next, fluorescently tagged versions of SOAP1, 2, 3, 4 and 5 were expressed in various combinations and ER co-localization of all proteins examined was observed (**FIGURES 35A-35F**, **FIGURE 52A-52F**). Detection of high quantities of MA-3-GlcA in leaves simultaneously expressing all fusion proteins demonstrated that the fluorescent signals observed originated from functional, and hence properly folded and subcellularly localized proteins (**FIGURES 53**). Also observed using fluorescence resonance energy transfer (FRET) was that SOAP1-5 proteins are located in proximity to each other and therefore possibly interact (**FIGURES 54A-54B**). Accurate colocalization of studied proteins could explain the high efficiency of performed reactions by decreasing the diffusion of the intermediates.

[001019] *Summary*: These studies showed that at least enzymes SOAP1-5 are localized in the ER in close proximity to each other.

[001020] ***General Conclusions from the Examples***

[001021] These studies on steroidal alkaloid, steroidal saponin, and triterpenoid saponin biosynthesis in tomato, potato, eggplant, and spinach resulted in discovery and characterization of genes that complete the biosynthetic pathway of complex triterpenoid saponins and elucidate the biosynthetic pathways of complex steroidal alkaloids and steroidal saponins. Discovery of the

unprecedented function of cellulose synthase like G proteins as glucuronosyltransferases of triterpenoid aglycones in seven plant species belonging to two distinct orders, proves that glycosylation of specialized metabolites is not exclusively performed by GTs belonging to 1 family Carbohydrate-Active enZYmes Database (CAZY). This finding will most likely trigger the discovery of additional functions of CSL enzymes related to other specialized metabolites (SMs). Additionally, the discovery of unique SOAP6 fucosyltransferase activity was demonstrated, as well as characterization of first BAHD acyltransferase (SOAP10) capable of acetylating sugar moieties of triterpenoid saponins. Similarly, GAME15 may in certain embodiments glycosylate steroidal alkaloids and steroidal saponins. In some embodiments, GAME15 glycoylation comprises transfer of a glucuronic acid (glucuronosyltransferase activity). Moreover, decreased expression of steroidal alkaloids and steroidal saponins provides a new approach to the increase in content of plant cholesterols, phytocholesterols, cholestenols, phytocholestenols, and phytosterols, which are useful as nutritional, cosmetic, and pharmaceutical agents,

[001022] The unexpected finding of CSLG proteins activity as triterpenoid glucuronosyltransferases shows that glycosylation of SMs is not exclusively executed by family 1 type UGTs. Hence, this report is likely to trigger the discovery of CSL enzymes functioning in modification of other classes of terpenoids and SMs, as well as in other classes of steroidal alkaloids and/or steroidal saponins. It moreover provides a fundamental example of how enzyme activity in one of the few principal plant processes, i.e., the cellulose synthesis machinery constructing cell walls, is 'hijacked' in order to produce a set of unrelated, defense specialized metabolites. Co-localization of SOAP5 at the spinach ER membrane with most other pathway proteins rather than the cytosol (in which UGTs localize) was part of evolving its new function. It suggests the importance of physical proximity between these enzymes for efficient triterpenoids production. This physical proximity was also reflected at the gene level as well, with respect to the *GAME* genes as a gene cluster in tomato. Apart from glucuronic acid, a yet undisclosed fucosyltransferase was identified acting on metabolites (i.e. SOAP6), as all plant fucosyltransferases to date were associated with fucosylation of cell wall polysaccharides and glycoproteins. While BAHD type acyltransferases are frequently associated with modification of SMs they were on no occasion reported to acylate triterpenoids as is demonstrated here for SOAP10 acetylating the C-28 fucose moiety of spinach triterpenoid saponins. With similarity to glucuronidation, the acylated fucose domain in triterpenoid saponins is important for their efficacy as therapeutic agents, e.g., adjuvanticity of the renowned QS-21, a potent saponin vaccine adjuvant from *Quillaja saponaria*.

[001023] This work delivers unparalleled strategy for heterologous engineering plants or yeast or other eukaryotic cell systems for sustainable production of high-value metabolites (e.g. glycyrrhizin,

soyasaponins, QS-21 adjuvant) for food and pharmaceutical industries by filling a gap in biosynthetic pathways of many glucuronide-type triterpenoid saponins. In addition, it supports engineering endogenous genes within a plant triterpenoid saponin biosynthetic pathway to either increase beneficial triterpenoid saponins, for example but not limited to glycyrrhizin in Chinese licorice, or decrease bitter or unwanted triterpenoid saponins, for example but not limited to their reduction in quinoa. Likewise, toxins and bitter tasting compounds are decreased by the reduction of unwanted steroidal alkaloids and/or steroidal saponins.

[001024] While certain features of genetically modified cells and uses thereof for producing steroidal alkaloids, steroidal saponins, and triterpenoid saponins have been illustrated and described herein, many modifications, substitutions, changes, and equivalents will now occur to those of ordinary skill in the art. The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying current knowledge, readily modify and/or adapt for various applications such specific embodiments without undue experimentation and without departing from the generic concept, and, therefore, such adaptations and modifications should and are intended to be comprehended within the meaning and range of equivalents of the disclosed embodiments. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation. The means, materials, and steps for carrying out various disclosed functions may take a variety of alternative forms without departing from the invention. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the disclosure herein.

CLAIMS

What is claimed is:

1. A genetically modified cell having increased expression of at least one heterologous gene compared to a corresponding unmodified cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said genetically modified cell comprises an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to a corresponding unmodified cell.

2. The genetically modified cell according to claim 1, wherein said genetically modified cell comprising an increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, further expresses at least one additional heterologous gene encoding an enzyme, said enzyme selected from the group consisting of a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, and an UDP-glucose 6-dehydrogenase 1, or any combination thereof.

3. The genetically modified cell according to any one of claims 1-2, wherein the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

4. The genetically modified cell according to any one of claims 1-2, wherein the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

5. The genetically modified cell according to any one of claims 2-4, wherein the nucleic acid sequence encoding said at least one additional heterologous gene encodes

   (a) a β-amyrin synthase, said nucleic acid sequence set forth in SEQ ID NO: 45; or

a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or

(b) a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or

(c) a glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or

(d) an acetyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or

(e) a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

6. The genetically modified cell according to any one of claims 2-4, wherein the amino acid sequence of said encoded at least one additional heterologous gene encodes

(a) a β-amyrin synthase, said amino acid sequence set forth in SEQ ID NO: 48; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 48; or

(b) a cytochrome P450, said amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or

(c) a glycosyl transferase, said amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or

(d) an acetyltransferase, said amino acid sequence set forth in SEQ ID NO: 64; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 64; or

(e) a UDP-glucose 6-dehydrogenase 1, said amino acid sequence set forth in SEQ

ID NO: 75; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 75; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

7. The genetically modified cell according to any one of claims 1-6, wherein said triterpenoid saponin or said derivative, metabolite, or biosynthetic intermediate thereof, comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof.

8. The genetically modified cell according to any one of claims 1-6,

(a) wherein said steroidal glycoalkaloid comprises an esculeoside, a dehydroesculeoside, alpha-tomatine, tomatine, dehydrotomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or

(b) wherein said steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof; or

(c) wherein said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), glycyrrhetinic acid 3-O-monoglucuronide (Compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or

(d) wherein the biosynthetic intermediate of said triterpenoid saponin comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or

(e) a combination thereof.

9. The genetically modified cell according to claim 1, having an altered content of a

phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

10. The genetically modified cell according to any one of claims 1-9, wherein said cell comprises a plant cell, a yeast cell, or an alga cell.

11. The genetically modified cell according to claim 10, wherein said plant cell comprises a cell from a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

12. The genetically modified cell according to claim 11, wherein:
    (a) when said plant cell comprises a cell from a plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus;
    (b) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus;
    (c) when said plant cell comprises a cell from a plant in the *Solanales* order, said plant is selected from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus;
    (d) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus;
    (e) when said plant cell comprises a cell from a plant in the *Malvales* order, said plant is selected from the *Theobroma* genus;
    (f) when said plant cell comprises a cell from a plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium*

genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus; or

(g) when said plant cell comprises a cell from a plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

13. The genetically modified cell according to claim 12, wherein:

(a) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group consisting of spinach, beetroot, and quinoa;

(b) when said plant cell comprises a cell from a plant in the *Solanales* order, said plant is selected from the group consisting of tomato, wild tomato, potato, wild potato, eggplant, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, ground cherry, tobacco, and bittersweet; or

(c) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

14. The genetically modified cell according to claim 13, wherein:

(a) the plant cell is from a tomato plant having:

(i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or dehydrotomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(b) the plant cell is from a potato plant having:

(i) an increased content of alpha-chaconine or a derivative thereof, a

metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(c) the plant cell is from an eggplant plant having:

(i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof;

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

15. The genetically modified cell according to any one of claims 10-14, wherein said plant cell comprises a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed or seed cell, a callus cell, a bract cell, or a flower cell.

16. The genetically modified cell according to any one of claims 10-15, wherein said plant cell is comprised in a plant or a portion thereof, said portion thereof comprising a plant leaf, a plant petiole, a plant stem or stalk, a plant root, a plant bud, a plant tuber, a plant

bean, a plant grain or kernel, a plant fruit, a plant nut, a plant legume, a plant seed, a plant bract, or a plant flower.

17. The genetically modified cell according to claim 10, wherein said yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus.

18. The genetically modified cell according to claim 10, wherein said alga is selected from a microalga, a multi-cellular alga, a cyanobacterium, a diatom, chlorophytes (green algae), rhodophytes (red algae), or phaeo-phytes (brown algae), a *Dunaliella*, a *Chlamydamonas*, or a *hematococus*.

19. A genetically modified plant comprising at least one cell having altered expression of at least a cellulose synthase like G (CSLG) gene compared to the expression of CSLG in a corresponding unmodified plant, and wherein the genetically modified plant has an altered content of at least one steroidal alkaloid, a derivative thereof a metabolite thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

20. The genetically modified plant of claim 19, wherein said CSLG gene comprises the endogenous CSLG gene.

21. The genetically modified plant of claim 20, wherein said endogenous CSLG gene comprises a mutation, said mutation comprising at least one or more point mutations, or an insertion, or a deletion, or any combination thereof, and

    (a) wherein said expressed CSLG enzyme has increased stability or increased activity or both and the altered content comprises an increased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant; or

    (b) wherein said expressed CSLG enzyme has decreased stability or decreased activity or both and the altered content comprises a decreased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a

biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

22. The genetically modified plant of claim 20, wherein:
    (a) said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression and said altered content comprises a reduced amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant; or
    (b) said endogenous CSLG gene is overexpressed and said altered content comprises an increased amount of the at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of the at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of the at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof compared to the corresponding unmodified plant.

23. The genetically modified plant of claim 22, wherein when said endogenous CSLG gene is selectively silenced, repressed, or has reduced expression, said cell further comprises at least one silencing molecule targeted to the polynucleotide encoding said CSLG gene, wherein the silencing molecule is selected from an RNA interference molecule or an antisense molecule, or wherein the silencing molecule is a component of a viral induced gene silencing system.

24. The genetically modified plant of claim 19, wherein said CSLG gene comprises a heterologous CSLG gene, and said altered expression comprises a *de novo* expression of said gene.

25. The genetically modified plant according to claim 24, wherein said expression comprises increased expression compared to a corresponding unmodified cell, and said altered content comprising increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one

steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

26. The genetically modified plant according to any one of claims 19-25, wherein the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40 , 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

27. The genetically modified plant according to any one of claims 19-25, wherein the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

28. The genetically modified plant according to any one of claims 19-27, wherein the altered expression of the at least one cellulose synthase like G (CSLG) gene is altered by introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, any combination thereof, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence.

29. The genetically modified plant according to claim 28, wherein the altered expression of the at least one CSLG gene is altered by introducing a silencing molecule targeted to the at least one CSLG gene and wherein:

    (a) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 30, the silencing molecule is set forth in SEQ ID NO: 42 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 42 or a complementary sequence thereof;

    (b) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 34, the silencing molecule is set forth in SEQ ID NO: 43 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 43 or a complementary sequence thereof;

    (c) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 38, the silencing

molecule is set forth in SEQ ID NO: 44 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 44 or a complementary sequence thereof;

(d) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 65 or SEQ ID NO: 93, the silencing molecule is set forth in SEQ ID NO: 106 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 65 or the nucleic acid sequence set forth in SEQ ID NO: 93; or

(e) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 95, the silencing molecule is set forth in SEQ ID NO: 107 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 95.

30. The genetically modified plant according to any one of claims 19-29, wherein said triterpenoid saponin, derivative thereof, metabolite thereof, or the biosynthetic intermediate thereof comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof.

31. The genetically modified plant according to any one of claims 19-31,

(a) wherein said steroidal glycoalkaloid comprises an esculeoside, a dehydroesculeoside, alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or

(b) wherein said steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof; or

(c) wherein said triterpenoid sapoinin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21

adjuvant, or any combination thereof; or

(d) wherein the biosynthetic intermediate of a triterpenoid saponin comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or

(e) a combination thereof.

32. The genetically modified plant according to claim 19, having an altered content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

33. The genetically modified plant according to any one of claims 19-32, wherein said plant is selected from the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order.

34. The genetically modified plant according to claim 33, wherein:

(a) when said plant comprises a plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus;

(b) when said plant comprises a plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus;

(c) when said plant comprises a plant in the *Solanales* order, said plant is selected from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus;

(d) when said plant comprises a plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus;

(e) when said plant comprises a plant in the *Malvales* order, said plant is selected from the *Theobroma* genus;

(f) when said plant comprises a plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium* genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus; or

(g) when said plant comprises a plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

35. The genetically modified plant according to claim 34, wherein:

(a) when said plant comprises a plant in the *Caryophyllales* order, said plant is selected from the group consisting of spinach, beetroot, and quinoa;

(b) when said plant comprises a plant in the *Solanales* order, said plant is selected from the group consisting of tomato, wild tomato, potato, wild potato, eggplant, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, ground cherry, tobacco, and bittersweet; or

(c) when said plant comprises a plant in the *Fabales* order, said plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

36. The genetically modified plant according to claim 35, wherein:

(a) the plant is a tomato plant having:

(i) an increased content of alpha-tomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or tomatidine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or dehydrotomatine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate

thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(b) the plant is a potato plant having:

(i) an increased content of alpha-chaconine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solanine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or

(c) the plant is an eggplant plant having:

(i) an increased content of alpha-solasonine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or alpha-solamargine or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a steroidal saponin or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof;

(ii) an altered content of a phytosterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

37. The genetically modified plant according to any one of claims 20-36, wherein said plant cell comprises a leaf cell, a petiole cell, a plant stem or stalk cell, a root cell, a bud cell, a tuber cell, a bean cell, a grain or kernel cell, a fruit cell, a nut cell, a legume cell, a seed

or seed cell, a callus cells, a bract cell, a callus cell, and a flower cell.

38. A method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin in a genetically modified cell, the method comprising:
    (a) introducing an at least one heterologous gene into said cell, said at least one heterologous gene encoding a cellulose synthase like G (CSLG) enzyme, wherein said heterologous gene is optionally comprised in a vector; and
    (b) expressing said at least one heterologous gene in said cell;

wherein said cell comprises an increased content of at least one steroidal alkaloid, at least one steroidal saponin, or at least one triterpenoid saponin compared to a corresponding unmodified cell.

39. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin of claim 38, said introducing further comprising introducing an at least one additional heterologous gene into said cell, said heterologous gene selected from the group consisting of the group encoding a saponin beta-amyrin synthase, a cytochrome P450, a glycosyltransferase, an acyltransferase, and a UDP-glucose 6-dehydrogenase I, or any combination thereof, wherein said at least one additional heterologous gene is optionally comprised in a vector; and further comprising expressing said at least one additional heterologous gene in said cell.

40. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin of any of claims 38-39, wherein
    (a) said at least one heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or
    (b) said at least one additional heterologous gene is operably linked to a promoter, a transcription termination sequence, or a combination thereof; or
    (c) a combination thereof of (a) and (b).

41. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-40, wherein said introducing comprises transforming said at least one cell with
    (a) said at least one heterologous gene or a polynucleotide sequence encoding said at least one heterologous gene, or the vector comprising said at least one heterologous gene; or
    (b) said at least one additional heterologous gene or a polynucleotide sequences encoding said at least one additional heterologous gene, or the vector comprising

said at least one additional heterologous gene; or

(c) a combination thereof of (a) and (b);

wherein said expressing comprises transient expression or constitutive expression.

42. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-41, wherein the nucleic acid sequence encoding said at least one heterologous CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40 , 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

43. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-42, wherein the amino acid sequence of said encoded CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

44. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-43, wherein the nucleic acid sequence encoding said at least one additional heterologous gene encodes

(a) a β-amyrin synthase, said nucleic acid sequence set forth in SEQ ID NO: 45; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 45; or

(b) a cytochrome P450, said nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 46, 51, or 53; or

(c) a glycosyl transferase, said nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NO: 55, 57, 59, or 61; or

(d) an acyltransferase, said nucleic acid sequence set forth in SEQ ID NO: 63; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 63; or

(e) a UDP-glucose 6-dehydrogenase 1, said nucleic acid sequence set forth in SEQ

ID NO: 74; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 74; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

45. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-43, wherein the amino acid sequence of said encoded at least one additional heterologous gene encodes

(a) a β-amyrin synthase, said amino acid sequence set forth in SEQ ID NO: 48; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 48; or

(b) a cytochrome P450, said amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 49, 52, or 54; or

(c) a glycosyl transferase, said amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in any one of SEQ ID NO: 56, 58, 60, or 62; or

(d) an acyltransferase, said amino acid sequence set forth in SEQ ID NO: 64; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 64; or

(e) a UDP-glucose 6-dehydrogenase 1, said amino acid sequence set forth in SEQ ID NO: 75; or a homolog thereof having at least 80% identity to and at least 80% coverage of, the amino acid sequence set forth in SEQ ID NO: 75; or

(f) any combination thereof of (a), (b), (c), (d), and (e).

46. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-45, wherein

(a) said triterpenoid saponin comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof; or

(b) said steroidal glycoalkaloid comprises an esculeoside, a dehydroesculeoside, alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or

(c) said steroidal saponin comprises uttroside B, a tomatoside, or any combination

thereof; or

(d) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or

(e) a combination thereof.

47. The method according to any one of claims 38-46, further comprising altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

48. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to any one of claims 38-47, wherein said cell comprises a plant cell, a yeast cell, or an alga cell.

49. The method of producing a steroidal alkaloid, a steroidal saponin, or a triterpenoid saponin according to claim 48, wherein

(a) said plant cell comprises a cell from a plant in the *Poales* order, the *Caryophyllales* order, the *Solanales* order, the *Fabales* order, the *Malvales* order, the *Apiales* order, the *Brassicales* order, the *Asparagales* order, the *Dioscoreales* order, or the *Liliales* order; or

(b) said yeast is selected from a *Saccharomyces* genus, a *Schizosaccharomyces* genus, a *Pichia* genus, a *Yarrowia* genus, a *Kluyveromyces* genus, or a *Candida* genus; or

(c) said alga is selected from a microalga, a multi-cellular alga, a cyanobacterium, a diatom, chlorophytes (green algae), rhodophytes (red algae), or phaeo-phytes

(brown algae), a *Dunaliella*, a *Chlamydamonas*, or a *hematococus*.

50. The method of producing a steroidal alkaloid, a steroidal saponoin, or a triterpenoid saponin according to claim 49, wherein:
    (a) when said plant cell comprises a cell from a plant in the *Poales* order, said plant is selected from the group of genera consisting of the *Oryza* genus, the *Hordeum* genus, the *Avena* genus, and the *Triticum* genus;
    (b) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group of genera consisting of the *Spinacia* genus, the *Chenopodium* genus, the *Beta* genus, the *Rheum* genus, the *Vaccaria* genus, the *Saponaria* genus, and the *Gypsophila* genus;
    (c) when said plant cell comprises a cell from a plant in the *Solanales* order, said plant is selected from the group of genera consisting of the *Solanum* genus, the *Capsicum* genus, the *Nicotiana* genus, the *Hyoscyamus* genus, the *Datura* genus, and the *Atropa* genus;
    (d) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group of genera consisting of the *Glycyrrhiza* genus, the *Medicago* genus, the *Glycine* genus, the *Lotus* genus, the *Cicer* genus, the *Phaseolus* genus, the *Pisum* genus, the *Arachis* genus, the *Lupinus* genus, and the *Acacia* genus;
    (e) when said plant cell comprises a cell from a plant in the *Malvales* order, said plant is selected from the *Theobroma* genus;
    (f) when said plant cell comprises a cell from a plant in the *Apiales* order, said plant is selected from the group of genera consisting of the *Daucus* genus, the *Apium* genus, the *Petroselinum* genus, the *Panax* genus, the *Bupleurum* genus, the *Hedera* genus, and the *Centella* genus; or
    (g) when said plant cell comprises a cell from a plant in the *Brassicales* order, said plant is selected from the group of genera consisting of the *Arabidopsis* genus, the *Brassica* genus, the *Capparis* genus, and the *Carica* genus.

51. The method according to claim 50, wherein:
    (a) when said plant cell comprises a cell from a plant in the *Caryophyllales* order, said plant is selected from the group consisting of spinach, beetroot, and quinoa; or
    (b) when said plant cell comprises a cell from a plant in the *Solanales* order, said plant is selected from the group consisting of tomato, wild tomato, potato, wild

potato, eggplant, pepper, bell pepper, cayenne pepper, chili pepper, pimiento, tabasco pepper, ground cherry, tobacco, and bittersweet; or

(c) when said plant cell comprises a cell from a plant in the *Fabales* order, said plant is selected from the group consisting of alfalfa, soy, *Lotus japonicus*, and licorice.

52. A method of reducing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or a plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising:

(a) transforming said at least one plant cell with at least one silencing molecule targeted to a nucleic acid gene sequence encoding a Cellulose Synthase Like G (CSLG) enzyme; or

(b) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence;

wherein expression of the gene encoding the CSLG enzyme is reduced in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell, wherein the plant comprising said genetically modified cell comprises reduced content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

53. The method of claim 52, wherein the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

54. The method of claim 52 or claim 53, wherein:

(a) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 30, the silencing molecule is set forth in SEQ ID NO: 42 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 42 or a complementary sequence thereof;

(b) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 34, the silencing molecule is set forth in SEQ ID NO: 43 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 43 or a complementary sequence thereof;

(c) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 38, the silencing molecule is set forth in SEQ ID NO: 44 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 44 or a complementary sequence thereof;

(d) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 65 or SEQ ID NO: 93, the silencing molecule is set forth in SEQ ID NO: 106 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 65 or the nucleic acid sequence set forth in SEQ ID NO: 93; or

(e) when the nucleic acid of the CSLG is set forth in SEQ ID NO: 95, the silencing molecule is set forth in SEQ ID NO: 107 or a complementary sequence thereof, or a homolog thereof having at least 80% identity to and at least 80% coverage of, the nucleic acid sequence set forth in SEQ ID NO: 95.

55. The method of any one of claims 52-54, wherein

(a) said steroidal alkaloid, said steroidal saponin, or said triterpenoid saponin comprises a toxin or a bitter tasting compound or having hormone mimicking properties, or a combination thereof; or

(b) said steroidal glycoalkaloid comprises an esculeoside, a dehydroesculeoside, alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or

(c) said steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof; or

(d) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or

(e) said intermediate comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or

(f) a combination thereof.

56. The method according to any one of claims 52-55, further comprising altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.

57. A method of increasing the content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, in at least one cell of a plant or plant part, the method comprising genetically modifying said at least one plant cell, said genetic modification comprising:

(a) mutagenizing at least one nucleic acid sequence encoding a Cellulose Synthase Like G (CSLG) enzyme, wherein the mutagenesis comprises introducing (1) one or more point mutations into the nucleic acid sequence, (2) deletions within the nucleic acid sequence, or (3) insertions within the nucleic acid, or (4) any

combination thereof, wherein said introducing comprising mutagenizing coding or non-coding sequence; and

(b) expressing said nucleic acid encoding said CSLG;

wherein the plant comprising said genetically modified cell comprises increased content of at least one steroidal alkaloid, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; of at least one steroidal saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or of at least one triterpenoid saponin, a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof, compared to the corresponding unmodified plant.

58. The method of claim 57, wherein the nucleic acid sequence encoding said CSLG gene is set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the nucleic acid sequence set forth in any one of SEQ ID NOS: 30, 32, 34, 36, 38, 40, 65, 80, 93, 95, 97, 99, 101, 103, or 105.

59. The method of claim 57, wherein the amino acid sequence of said CSLG enzyme is set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104; or a homolog thereof having at least 55% identity to and at least 55% coverage of, the amino acid sequence set forth in any one of SEQ ID NOS: 31, 33, 35, 37, 39, 41, 66, 81, 94, 96, 98, 100, 102, or 104.

60. The method of any one of claims 57-59, wherein

(a) said steroidal alkyloid, said steroidal saponin, or said triterpenoid saponin comprises a sweetener, foaming agent, emulsifier, preservative, anti-carcinogen, hypocholesterolemic agent, anti-inflammatory agent, anti-oxidant, biological adjuvant, anti-microbial agent, insecticidal agent, antifeedant, anti-fungal agent, or any combination thereof; or

(b) said steroidal glycoalkaloid comprises an esculeoside, a dehydroesculeoside, alpha-tomatine, tomatine, alpha-chaconine, alpha-solanine, alpha-solasonine, alpha-solmargine, or any combination thereof; or

(c) wherein said steroidal saponin comprises uttroside B, a tomatoside, or any combination thereof; or

(d) said triterpenoid saponin comprises medicagenic acid 3-O-glucuronide (MA-3-GlcA) (Compound 6), Compound 7, Compound 8, Compound 9, Compound 10, Compound 11, glycyrrhizin (Compound 14), Glycyrrhetinic acid 3-O-

monoglucuronide (compound 15), bayogenin (Compound 25), bayogenin-hexA-hex-hex (Compound 31), serjanic acid (Compound 26), serjanic acid-hexA-hex (Compound 32), soyasapogenol A (Compound 29), soyasapogenol B (Compound 30), soyasapogenol A–hexA-hex-pent (Compound 34), soyasaponin VI (Compound 35), betavulgaroside IV (Compound 33), hederagenin-3GlcA, gypsogenin-3GlcA, gypsogenic acid-3GlcA, or a QS-21 adjuvant, or any combination thereof; or

(e) said intermediate comprises Compound 1, Compound 2, medicagenic acid (Compound 5), oleanolic acid (Compound 3), oleanolic acid-3GlcA, augustic acid (Compound 4), augustic acid-3GlcA, or glycyrrhetinic acid (Compound 13), or any combination thereof; or

(f) a combination thereof.

61. The method of any one of claim 57-60, wherein

(a) expression of the gene encoding the CSLG enzyme is increased in the genetically modified plant cell compared to its expression in a corresponding unmodified plant cell; or

(b) said encoded CSLG enzyme has increased activity in the genetically modified plant cell compared to its activity in a corresponding unmodified plant cell; or

(c) said encoded CSLG enzyme has increased stability in the genetically modified plant cell compared to its stability in a corresponding unmodified plant cell; or

(d) any combination thereof of (a), (b), and (c).

62. The method according to any one of claims 57-61, further comprising altering the content of a phytosterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a phytocholesterol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; a cholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof; or a phytocholestenol, or a derivative thereof, a metabolite thereof, or a biosynthetic intermediate thereof.