

TERPENE SYNTHASES FROM YLANG YLANG
(CANANGA ODORATA VAR. FRUTICOSA)

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is related and claims priority to U.S. provisional patent application Serial No. 62/067,191 filed 22 October 2014. This application is incorporated herein by reference.

SEQUENCE SUBMISSION

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is entitled 2577241SequencListing.txt, was created on 9 October 2015 and is 287 kb in size. The information in the electronic format of the Sequence Listing is part of the present application and is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0003] The present invention relates to the field of plant molecular biology. More particularly, the present invention relates to the isolation of nucleic acids encoding terpene synthases (TPSs), including a novel, multifunctional TPS identified herein as CoTPS2.

[0004] The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated by reference, and for convenience are referenced in the following text by author and date and are listed alphabetically by author in the appended bibliography.

[0005] Plants emit a large group of phytogetic volatile organic compounds (VOCs) for their defense against pathogens, parasites and herbivores and for attracting pollinators (Kessler and Baldwin, 2001; Dudareva et al., 2013). VOCs are synthesized in all plant organs such as flowers, stems, leaves, roots, fruits and seeds, but the quantity and diversity of VOCs change in response to environmental stimuli (Dudareva et al., 2013). VOCs are major components of floral scent in a wide range of flowers (Knudsen et al., 1993). Over 1,700 floral VOCs have been identified from 90 different plants, and they are assumed to have functions in both attraction of pollinators and defence against pathogens (Knudsen et al., 2006; Muhlemann et al., 2014). Given the role of VOCs, the production and emission of VOCs are highly regulated spatially and developmentally. Floral VOCs are mainly composed of terpenoids, phenylpropanoids/benzenoids and volatile fatty acid derivatives which are derived from different biosynthetic routes in plants (Muhlemann et al., 2014). Terpenoids, also referred to as isoprenoids, are the largest and most diverse class of VOCs in plants (Dudareva et al., 2013). Terpenes are

synthesized from two distinct and compartmentally separated pathways, the mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP) pathways (McGarvey and Croteau, 1995). Phenylpropanoids and benzenoids class of metabolites are primarily derived from the carbon skeleton of phenylalanine which is produced by the shikimate pathway (Vogt, 2010; Orlova et al., 2006).

[0006] Terpene synthases (TPSs) are responsible for generating the immense diversity in terpenes produced by plants (McGarvey and Croteau, 1995). Many TPSs have the ability to synthesize multiple products from a single prenyl diphosphate substrate (Degenhardt et al., 2009). Based on the sequence relatedness and functional assessment, the TPS gene family has been divided into seven subfamilies designated TPS-a through TPS-g (Bohlmann et al., 1998; Lee and Chappell, 2008; Martin et al., 2010). TPS-a subfamily typically contains angiosperm-specific sesqui-TPSs, whereas angiosperm mono-TPSs form the TPS-b subfamily. The TPS-b subfamily contains the arginine-tryptophan motif, R(R)X₈W (SEQ ID NO:1) which plays a role in the RR-dependent isomerization of GPP (Martin et al., 2010). Another angiosperm mono-TPS subfamily, the TPS-g contains members of mono-TPSs that lack the R(R)X₈W (SEQ ID NO:1) motif characteristic. These TPSs produce acyclic monoterpenes that contribute to the floral VOCs (Dudareva et al., 2013). TPS-c and TPS-e subfamilies consist of angiosperm di-TPSs responsible for gibberellic acid (GA) biosynthesis namely copalyl diphosphate synthases (CPS) and kaurene synthases (KS). The different mono-, sesqui-, and di-TPS genes for synthesis of conifer-specialized terpenes belong to the gymnosperm-specific TPS-d subfamily (Martin et al., 2004). TPS-f includes the monoterpene linalool synthase of the genus *Clarkia* (Dudareva et al., 1996).

[0007] *Cananga odorata*, commonly called ylang ylang is a tropical evergreen tree of the *Annonaceae* family that produces fragrant flowers and is widely cultivated throughout Southeast Asia. Essential oils obtained by steam distillation from mature fresh ylang ylang flowers are used in the cosmetic industry as major components of perfumes and fragrances, in the food industry as ingredients of aromas and flavours, and in the pharmaceutical industry as active components of antibacterials and in aromatherapy (Benini et al., 2010; Burdock and Carabin, 2008; Gaydou et al., 1986). The chemical composition of floral VOCs produced by ylang ylang varieties has been previously reported (Gaydou et al., 1986; Benini et al., 2010, 2012; Brokl et al., 2013). These papers show the presence of volatile terpenes, benzenoid and phenylpropanoids in floral VOCs. Gaydou et al., described the composition of essential oils of ylang ylang flowers originating from Madagascar (*Cananga odorata* Hook Fil. et Thomson forma genuina). These

authors found that the primary component was the monoterpene linalool (19%), and the other major compounds were two sesquiterpenes, β -caryophyllene (10.7%) and germacrene D (10.3%) (Gaydou et al., 1986). Additionally, this variety of ylang ylang from Madagascar contained more than 20% of other aromatic compounds such as p-methylanisole, benzyl benzoate, methyl benzoate and benzyl salicylate (Gaydou et al., 1986). *Cananga odorata* var. *fruticosa* or dwarf ylang ylang is another variety which is popularly grown in Southeast Asia as a small and compact shrub with highly scented flowers. Its essential oil is also used in the perfume industry. Despite the economic and social importance of this species, the biosynthetic pathways leading to the production of the floral scent of ylang ylang have not been fully understood.

SUMMARY OF THE INVENTION

[0008] The present invention relates to the field of plant molecular biology. More particularly, the present invention relates to the isolation of nucleic acids encoding terpene synthases (TPSs), including a novel, multifunctional TPS identified herein as CoTPS2.

[0009] As described herein, the chemical composition of floral VOC was determined at 4 different stages of flower formation and performed RNA-seq on mature yellow flowers of *C. odorata* var. *fruticosa* where the production of floral VOCs is at the maximum. Terpenes formed the bulk of floral VOCs. The transcriptome data revealed 16 TPS transcripts from dwarf ylang ylang flowers out of which 4 were functionally characterized. Two TPSs were identified as mono-TPSs, CoTPS1 and CoTPS4, which catalyze the formation of multiproduct β -thujene/sabinene/ β -pinene/ α -terpinene and geraniol, respectively. The other two CoTPS2 and CoTPS3 were ascertained as sesqui-TPSs. CoTPS3 was shown to catalyze the formation of α -bergamotene. CoTPS2 was found to be a multifunctional and novel TPS which could synthesize three sesquiterpene compounds, β -ylangene/ β -copaene/ β -cubebene from the farnesyl pyrophosphate (FPP) substrate.

[0010] Thus, in one aspect, the present invention provides an isolated nucleic acid encoding a CoTPS protein. In one embodiment, the nucleic acid further comprises a plant operable promoter operably linked to the coding sequence. In some embodiments, the CoTPS protein is CoTPS2 protein comprising the amino acid sequence set forth in SEQ ID NO:10. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:9. In a further embodiment, the nucleic acid encodes a variant CoTPS2 protein. In one embodiment, the variant CoTPS2 protein has at least 90% sequence identity with the CoTPS2 protein while having the activity of the CoTPS2 protein. In another embodiment, the variant CoTPS2 protein

has one or more amino acid changes in the amino acid sequence of the CoTPS2 protein while having the activity of the CoTPS2 protein.

[0011] In other embodiments, the CoTPS protein is CoTPS3 protein comprising the amino acid sequence set forth in SEQ ID NO:6. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:5. In a further embodiment, the nucleic acid encodes a variant CoTPS3 protein. In one embodiment, the variant CoTPS3 protein has at least 90% sequence identity with the CoTPS3 protein while having the activity of the CoTPS3 protein. In another embodiment, the variant CoTPS3 protein has one or more amino acid changes in the amino acid sequence of the CoTPS3 protein while having the activity of the CoTPS3 protein.

[0012] In some embodiments, the CoTPS protein is CoTPS1 protein comprising the amino acid sequence set forth in SEQ ID NO:8. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:7. In a further embodiment, the nucleic acid encodes a variant CoTPS1 protein. In one embodiment, the variant CoTPS1 protein has at least 90% sequence identity with the CoTPS1 protein while having the activity of the CoTPS1 protein. In another embodiment, the variant CoTPS1 protein has one or more amino acid changes in the amino acid sequence of the CoTPS1 protein while having the activity of the CoTPS1 protein.

[0013] In other embodiments, the CoTPS protein is CoTPS4 protein comprising the amino acid sequence set forth in SEQ ID NO:12. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:11. In a further embodiment, the nucleic acid encodes a variant CoTPS4 protein. In one embodiment, the variant CoTPS4 protein has at least 90% sequence identity with the CoTPS4 protein while having the activity of the CoTPS4 protein. In another embodiment, the variant CoTPS4 protein has one or more amino acid changes in the amino acid sequence of the CoTPS4 protein while having the activity of the CoTPS4 protein.

[0014] In a second aspect, the present invention provides a construct or vector comprising an isolated nucleic acid as described herein. In one embodiment, the construct or vector is an expression construct or vector. In another embodiment, the construct or vector further comprises a selectable marker. In a further embodiment, the construct or vector comprises a recombination marker free system. In some embodiments, the recombination marker free system is a Cre-lox recombination marker free system, a Zinc finger marker free system, a TALE nucleases marker free system or a CRISPR-Cas marker free system.

[0015] In a third aspect, the present invention provides a transgenic plant comprising a nucleic acid, construct or vector described herein. In one embodiment, the transgenic plant may be any plant species. In another embodiment, the transgenic plant may be a member of the *Cananga* genus. In an additional embodiment, the transgenic plant is a *Cananga odorata* plant. In a further embodiment, the transgenic plant may be *Cananga odorata* var. *fruticosa* (dwarf ylang ylang). In one embodiment, the transgenic *Cananga* plant or *Cananga odorata* plant or *Cananga odorata* var. *fruticosa* plant overexpresses a CoTPS protein compared to non-transgenic plants. In other embodiments, the transgenic plant expresses a lower amount of a CoTPS protein compared to a non-transgenic plant.

[0016] In a fourth aspect, the present invention provides a method for manipulating synthesis of a terpene or sesquiterpene in a transgenic plant to alter fragrance/flavor characteristics and/or plant pathogen interactions and/or insect interactions. In one embodiment, the transgenic plant may be any plant species. In another embodiment, the transgenic plant may be a *Cananga odorata* plant. In a further embodiment, the transgenic plant may be *Cananga odorata* var. *fruticosa* (dwarf ylang ylang). In one embodiment, the synthesis of the sesquiterpenes β -ylangene, β -copaene and β -cubebene is manipulated by overexpressing the CoTPS2 protein in the transgenic plant compared to a non-transgenic plant. In another embodiment, the synthesis of the sesquiterpene α -bergamotene is manipulated by overexpressing the CoTPS3 protein in the transgenic plant compared to a non-transgenic plant. In an additional embodiment, the synthesis of the terpene β -thujene/sabinene/ β -pinene/ α -terpinene is manipulated by overexpressing the CoTPS1 protein in the transgenic plant compared to a non-transgenic plant. In a further embodiment, the synthesis of the terpene geraniol is manipulated by overexpressing the CoTPS4 protein in the transgenic plant compared to a non-transgenic plant. In some embodiments, the synthesis of the described terpenes or sesquiterpenes is manipulated by reducing expression of the corresponding CoTPS protein in the transgenic plant compared to a non-transgenic plant.

[0017] In a fifth aspect, the present invention provides a method of preparing at least one terpene or sesquiterpene comprising the steps of (a) culturing a cell which has been genetically modified with a polynucleotide of the invention to provide CoTPS activity and (b) separating the at least one terpene or sesquiterpene produced. In one embodiment, the cells may be any cell type that can be grown in culture. In another embodiment, the cells may be bacteria or yeast cells for producing the terpene or sesquiterpene. In one embodiment, the at least one sesquiterpene is β -ylangene, β -copaene and/or β -cubebene and the CoTPS activity is CoTPS2 activity. In another embodiment, the at least one sesquiterpene is α -bergamotene and the CoTPS

activity is CoTPS3 activity. In an additional embodiment, the at least one terpene is β -thujene/sabinene/ β -pinene/ α -terpinene and the CoTPS activity is CoTPS1 activity. In a further embodiment, the at least one terpene is geraniol and the CoTPS activity is CoTPS4 activity. In one embodiment, the cell contains a metabolic pathway for producing farnesyl pyrophosphate which provides the farnesyl pyrophosphate that is the substrate for CoTPS2 or CoTPS3. In another embodiment, the cell has been transformed with a nucleic acid encoding farnesyl pyrophosphate synthase to produce farnesyl pyrophosphate which provides the farnesyl pyrophosphate that is the substrate for CoTPS2 or CoTPS3. In one embodiment, the cell contains a metabolic pathway for producing geranyl diphosphate which provides the geranyl diphosphate that is the substrate for CoTPS1 or CoTPS4. In another embodiment, the cell has been transformed with a nucleic acid encoding geranyl diphosphate synthase to produce geranyl diphosphate which provides the geranyl diphosphate that is the substrate for CoTPS1 or CoTPS4.

BRIEF DESCRIPTION OF THE FIGURES

[0018] Figures 1A and 1B show the compositional variation of ylang ylang essential oils during flower development. **Figure 1A:** The photograph of ylang ylang flowers showing the development of newly emerged flower bud to fully open yellow flower. I, undeveloped small flower; II, mature green flower; III, fully mature yellow flower. Scale bar, 1 cm. **Figure 1B:** GC trace of essential oils from floral buds, undeveloped small flowers (I), mature green flowers (II), and fully mature yellow flowers (III). The arrows indicate the retention time of camphor (10 $\mu\text{g}/\mu\text{l}$) as an internal standard. The peaks numbered in GC traces were identical to those listed in Table 2. These are magnified images of the corresponding zones indicated by a dotted line in Figure 9.

[0019] Figure 2 shows the GO terms for the top 1,000 highly expressed transcripts in dwarf ylang ylang flowers.

[0020] Figures 3A-3D show the qRT-PCR analyses of different biosynthetic pathway genes. Expression of genes involved in MEP (**Figure 3A**), MVA (**Figure 3B**), shikimate (**Figure 3C**), and phenyl propanoid (**Figure 3D**) were examined from leaves (L), buds (B) and three different stages of flower development, undeveloped small flower (I), mature green flower (II) and fully mature yellow flower (III) by qRT-PCR. MEP, 2-C-methyl-D-erythritol 4-phosphate; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMS, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-

(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase; MCS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4- hydroxy-3-methylbut-2-en-1-yl diphosphate synthase; IDI, isopentenyl pyrophosphate isomerase; MVA, mavalonate; AACT, acetyl-CoA acetyltransferase; HMGS, hydroxymethylglutaryl- CoA synthase; HMGR, hydroxymethylglutaryl- CoA reductase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MPDC, mevalonate diphosphate decarboxylase; DAHPS, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; DHQS, 3-dehydroquinate synthase; SDH, shikimate dehydrogenase; SK, Shikimate kinase; CS, chorismate synthase; PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; CCR, cinnamoyl-CoA reductase; C3H, p-coumarate-3-hydroxylase; CCOMT, caffeoyl-CoA 3-O-methyltransferase; CAD, cinnamyl alcohol dehydrogenase; CVOMT, chavicol O-methyltransferase; COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase.

[0021] Figures 4A and 4B show the phylogenetic analysis and alignment of TPSs from dwarf ylang ylang. **Figure 4A:** The maximum likelihood tree was drawn by MEGA 6 program from an alignment of full-length CoTPSs with other plant TPSs. AaADS, *Artemisia annua* (Aa) amorpho-411-diene synthase(AFA34434); AaFS, β -farnesene synthase (Q9FX7); AaQHS1, β -caryophyllene synthase (AAL79181); AaQH5, linalool synthase (AAF13356); Ag10, *Abies grandis* (Ag) 4S-limonene synthase (AAB70907); Ag2, myrcene synthase (AAB71084); Ag4, δ -selinene synthase (AAC05727); Ag3, pinene synthase (AAB71085); Ag9, terpinolene synthase (AAF61454); AmNES/LIS-1, *Antirrhium majus* (Am) nerolidol/linalool synthase1 (ABR24417); AmMS, myrcene synthase (AAO41727); AtCPS1, *Arabidopsis thaliana* (At) copalyl diphosphate synthase (NP_192187); AtGA2, kaurene synthase; AtTPS10 (AAC39443), myrcene/ocimene synthase (AAG09310); AtTPS14, linalool synthase (NP176361); CbLIS, *Clarkia breweri* S-linalool synthase (AAC49395); CILS1, *Citrus limon* limonene synthase 1 (AAM53944); CmCPS, *Cucurbits maxima* copalyl diphosphate synthase (AAD04292); CrGES, *Catharanthus roseus* geraniol synthase (AFD64744); FaNES2, *Fragaria x ananassa* nerolidol synthase (CAD57081); LsLTC1, *Lactuca sativus* germacrene A synthase (AAM11626); MpFS, *Mentha piperita* β -farnesene synthase (AAB95209); MsLS, *Mentha spicata* 4S-limonene synthase (AAC37366); ObGES, *Ocimum basilicum* geraniol synthase (AAR11765); OeGES1, *Olea europaea* geraniol synthase 1 (AFI47926); PaTPS-Bis, *Picea abies* α -bisabolene synthase (AAS47689); PcLS, *Perilla citriodora* limonene synthase (AAG31435); PfLS, *Perilla frutescens* linalool synthase (AAL38029); SaSSy, *Santalum album* santalene/bergamotene synthase (ADO87000); ScGAS, *Solidago canadensis* (Sc) germacrene A synthase (CAC36896);

ScGDS, germacrene D synthase (AAR31145); ShSBS, *Solanum habrochaites* santalene/bergamotene synthase (B8XA41); SlSBS, *Solanum lycopersicum* (Sl) santalene and bergamotene synthase (XP004244438); SlCPS, copalyl diphosphate synthase (BAA84918); SlGCS, germacrene C synthase (AAC39432); santalene/bergamotene synthase (BAA84918); Sl TPS38, Sl terpene synthase 38 (AEP82768); SoBPS, *Salvia officinalis* (So) bornyl diphosphate synthase (AAC26017); SoCS, 1,8-cineole synthase (AAC26016); SoSS, sabinene synthase (AAC26018); SrCPS, *Stevia rebaudiana* (Sr) copalyl pyrophosphate synthase (AAB87091); SrKS, kaurene synthase (AAD34294); VvPNGer, *Vitis vinifera* geraniol synthase (ADR74218); ZmTPS1, *Zea mays* terpene synthase 1 (AAO18435). Seven TPS subfamilies, a to g are delimited by dashed lines as based on the taxonomic distribution of the TPS families (Chen et al., 2011). **Figure 4B:** Comparison of deduced amino acid sequences of dwarf ylang ylang TPSs. The deduced amino acid sequences of *CoTPSs* were aligned using CLUSTAL W method. The Asp-rich domain, DDXXD (SEQ ID NO:2), the R(R)X₈W (SEQ ID NO:1) motif, and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) motif, which are highly conserved in plant TPSs and required for TPS activity, are indicated on the sequences. The arrowheads denote the predicted cleavage sites of plastidial transit peptides. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment. The deduced amino acid sequences for the *CoTPSs* are: *CoTPS3* – SEQ ID NO:6; *CoTPS1* - SEQ ID NO:8; *CoTPS2* – SEQ ID NO:10; and *CoTPS4* – SEQ ID NO:12. Nucleic acid sequences encoding these amino acid sequences are set forth in SEQ ID NOs:5, 7, 9 and 11, respectively.

[0022] Figure 5 shows the subcellular localization of *CoTPSs*. YFP-fused *CoTPSs* (*CoTPS1*-YFP, *CoTPS2*-YFP, *CoTPS3*-YFP, and *CoTPS4*-YFP) were transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration and visualized 3 dpi (days post-infiltration) using YFP channel of a confocal microscope. Auto, chlorophyll autofluorescence; YFP, YFP channel image; Light, light microscope image; Merged, merged image between Auto and YFP. Scale bars, 50 μ m.

[0023] Figure 6 shows transcript levels of dwarf ylang ylang *TPS* genes, *CoTPS1*, *CoTPS2*, *CoTPS3*, and *CoTPS4* during flower development. Total RNAs were isolated from leaves (L), floral buds (B) and three different stages of flower development, undeveloped small flower (I), mature green flower (II) and fully mature yellow flower (III) and used as the templates for qRT-PCR. Amplification of *Actin* mRNA was used as internal control.

[0024] Figures 7A-7D show *in vitro* enzymatic assays of recombinant CoTPSs. *In vitro* enzyme assays using recombinant His-tagged CoTPS1 (**Figure 7A**), CoTPS2 (**Figure 7B**), CoTPS3 (**Figure 7C**), or CoTPS4 (**Figure 7D**) protein using GPP or FPP as substrate. The reaction products were analysed by GC-MS. The peaks marked with an arrow in the GC traces were identified by the mass spectra reference library. Mass spectra for the peaks formed with FPP or GPP are shown on the right side of the figure with the references. *m/z*, mass-to-charge ratio.

[0025] Figures 8A and 8B show *in vivo* characterization of CoTPS2 and CoTPS3. YFP-fused CoTPS2 (**Figure 8A**) or CoTPS3 (**Figure 8B**) with or without HMGR was transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi by GC-MS. Numbered peaks were identified by the mass spectra reference library and the mass spectra of compounds were shown on the right side. The expression of HMGR alone in each figure was used as a control. The asterisk indicates a nonspecific peak derived from the expression of HMGR in *N. benthamiana* leaves

[0026] Figure 9 shows the compositional variation of dwarf ylang ylang essential oils during flower development. B, floral buds; I, undeveloped small flowers; II, mature green flowers; III, fully mature yellow flowers.

[0027] Figures 10A and 10B show total ion chromatograms of essential oils from dwarf ylang ylang flowers. **Figure 10A**: GC traces showing no difference between fresh and dried flowers. **Figure 10B**: GC traces of flowers and leaves from dwarf ylang ylang. 1, α -pinene; 2, Elixene; 3, β -elemene; 4, β -caryophyllene; 5, γ -muurolene; 6, Humulene; 7, Germacrene D; 8, γ -elemene; 9 and 10, phytol.

[0028] Figure 11 shows variation of four major terpenes during flower development. Bud, floral buds; I, undeveloped small flowers; II, mature green flowers; III, fully mature yellow flowers.

[0029] Figure 12 shows the total ion chromatograms of essential oils from dwarf ylang ylang flowers sampled at day and night.

[0030] Figure 13 shows the quality of deep sequencing. The sequence quality was evaluated by FastQC.

[0031] Figure 14A-14D shows the alignment of deduced amino acid sequences of representative genes involved in biosynthetic pathways for VOCs. **Figure 14A**: Alignment of deduced amino acid sequences of two representative genes, DXR and CMK, involved in MEP pathway. DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMK, 4-(cytidine 5'-

diphospho)-2-C-methyl-D-erythritol kinase. Co, *Cananga odorata*; Cr, *Catharanthus roseus*; Eu, *Eucommia ulmoides*; Lj, *Lonicera japonica*; Rc, *Ricinus communis*; Sm, *Salvia miltiorrhiza*; Zm, *Zea mays*. The CoDXR sequence is SEQ ID NO:111, and the CoCK sequence is SEQ ID NO:115. Accession numbers: RcDXR (XP_002511399; SEQ ID NO:112), EuDXR (AFU93070; SEQ ID NO:113), ZmDXR (NP_001105139; SEQ ID NO:114), CrCMK (ABI35992; SEQ ID NO:116), SmCMK (ABP96842; SEQ ID NO:117), LjCMK (AGE10581; SEQ ID NO:118).

Figure 14B: Alignment of deduced amino acid sequences of two representative genes, HMGS and PMK, involved in MVA pathway. HMGS, hydroxymethylglutaryl-CoA synthase; PMK, phosphomevalonate kinase. Ae, *Arnebia euchroma*; Co, *Cananga odorata*; Gs, *Glycine soja*; Nt, *Narcissus tazetta*; Pn, *Panax notoginseng*; Zm, *Zea mays*. The CoHMGS sequence is SEQ ID NO:119, and the CoPMK sequence is SEQ ID NO:123. Accession number of the orthologous genes used in the alignment: NtHMGS (AHF81872; SEQ ID NO:120), GsHMGS (KHN14128; SEQ ID NO:121), PnHMGS (AIK21781; SEQ ID NO:122), PnPMK (AIK21784; SEQ ID NO:124), ZmPMK (NP_001149345; SEQ ID NO:125), AePMK (ABY27562; SEQ ID NO:126).

Figure 14C: Alignment of deduced amino acid sequences of two representative genes, DAHPS and CS, involved in shikimate pathway. DAHPS, 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; CS, chorismate synthase. Co, *Cananga odorata*; Hb, *Hevea brasiliensis*; Mt, *Medicago truncatula*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Vv, *Vitis vinifera*. The CoDAHPS sequence is SEQ ID NO:127, and the CoCS sequence is SEQ ID NO:131. Accession number of the orthologous genes used in the alignment: VvDAHPS (NP_001268127; SEQ ID NO:128), MtDAHPS (XP_003615152; SEQ ID NO:129), HbDAHPS (AFY09700; SEQ ID NO:130), RcCS (XP_002529571; SEQ ID NO:132), HbCS (ADR70879; SEQ ID NO:133), PtCS (XP_002315301; SEQ ID NO:134).

Figure 14D: Alignment of deduced amino acid sequences of two representative genes, PAL and C4H, involved in phenyl propanoid pathway. PAL, phenylalanine ammonia lyase; C4H, cinnamate-4-hydroxylase. Co, *Cananga odorata*; Ga, *Gossypium arboreum*; Gm, *Glycine max*; Mb, *Musa balbisiana*; Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Tc, *Theobroma cacao*. The CoPAL sequence is SEQ ID NO:135, and the CoC4H sequence is SEQ ID NO:139. Accession number of the orthologous genes used in the alignment: PtPAL (ACC63889; SEQ ID NO:136), MbPAL (BAG70992; SEQ ID NO:137), RcPAL (XP_002531677; SEQ ID NO:138), GmC4H (ACR44227; SEQ ID NO:140), TcC4H (XP_007011365; SEQ ID NO:141), GaC4H (AAG10197; SEQ ID NO:142).

[0032] Figure 15 shows a phylogenetic analysis of DXSs from dwarf ylang ylang. The maximum likelihood tree was drawn by MEGA 6 program from an alignment of full-length

CoDXSs (1-deoxy-D-xylulose 5- phosphate synthase) with other plant DXSs. *At*, *Arabidopsis thaliana*; *Os*, *Oryza sativa*; *Pa*, *Picea abies*; *Pt*, *Populus trichocarpa*; *Sm*, *Salvia miltiorrhiza*.

[0033] Figure 16 shows the comparison of deduced amino acid sequences of four CoDXS small gene family. The thiamin diphosphate-binding site and the pyridine binding DRAG domain are indicated by the open box and the horizontal line, respectively. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment. The sequences are as follows: CoDXS1: SEQ ID NO:143; CoDXS2: SEQ ID NO:144; CoDXS3: SEQ ID NO:145; and CoDXS4: SEQ ID NO:146.

[0034] Figures 17A-17D shows the alignment of deduced amino acid sequences of CoTPSs and other plant TPSs. **Figure 17A:** Alignment of deduced amino acid sequences of CoTPS1 (SEQ ID NO:8) and other plant TPSs. PsSS, *Picea sitchensis* sabinene synthase (ADU85930; SEQ ID NO:149); SoSS, *Salvia officinalis* sabinene synthase (AAC26018; SEQ ID NO:148); Mg17, *Magnolia grandiflora* α -terpineol synthase (B3TPQ7; SEQ ID NO:147). The conserved motifs DDXXD (SEQ ID NO:2) and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) are marked. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment. **Figure 17B:** Alignment of deduced amino acid sequences of CoTPS2 (SEQ ID NO:10) and other plant TPSs. RcSeTPS1, *Ricinus communis* α -copaene synthase (B9S9Z3; SEQ ID NO:150); Mg25, *Magnolia grandiflora* β -cubebene synthase (B3TPQ6; SEQ ID NO:151); HaCS, *Helianthus annuus* α -copaene synthase (Q4U3F6; SEQ ID NO:152). The conserved motifs DDXXD (SEQ ID NO:2) and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) are marked. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment. **Figure 17C:** Alignment of deduced amino acid sequences of CoTPS3 (SEQ ID NO:6) and other plant TPSs. LcTOS, *Litsea cubeba* trans-ocimene synthase (AEJ91554; SEQ ID NO:153); LaBERS, *Lavandula angustifolia* α -bergamotene synthase (Q2XSC4; SEQ ID NO:154); LdTPS7, *Lippia dulcis* α -bergamotene synthase (J7LQ09; SEQ ID NO:155). The conserved motifs DDXXD (SEQ ID NO:2) and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) are marked. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment. **Figure 17D:** Alignment of deduced amino acid sequences of CoTPS4 (SEQ ID NO:12) and other plant TPSs. ObGES, *Ocimum basilicum*

geraniol synthase (AAR11765; SEQ ID NO:157); OeGES1, *Olea europaea* geraniol synthase 1 (AFI47926; SEQ ID NO:156); CrGES, *Catharanthus roseus* geraniol synthase (AFD64744; SEQ ID NO:158). The conserved motifs DDXXD (SEQ ID NO:2) and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) are marked. Completely conserved residues are shaded in dark gray, and identical residues are shaded in gray, and similar residues are shaded in light gray. Dashes indicate gaps introduced to maximize sequence alignment.

[0035] Figure 18 shows a SDS-PAGE of recombinant His-tagged CoTPS1, CoTPS2, CoTPS3, or CoTPS4 protein. M; Precision Plus Protein all blue standards (Bio-rad), kDa; kilodalton, -; minus IPTG, +; plus IPTG, PP; Purified protein.

[0036] Figure 19 shows *in vitro* enzymatic assay of recombinant 6His-tagged CoTPS4s using GPP. The reaction products were analysed by GC-MS. The peak was identified by the mass spectra reference library and comparison of retention time with those of authentic standard (geraniol standard). Mass spectra for the peak formed with GPP are shown on the right side of the figure with the references. *m/z*, mass-to-charge ratio. Control, control assay using a heat-activated recombinant protein.

[0037] Figure 20 shows transient expression of CoTPS2-YFP or YFP in *N. benthamiana*. CoTPS2-YFP or YFP was transiently expressed in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration. The compounds were analysed 3 dpi (days post-infiltration) by GC-MS.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The present invention relates to the field of plant molecular biology. More particularly, the present invention relates to the isolation of nucleic acids encoding terpene synthases (TPSs), including a novel, multifunctional TPS identified herein as CoTPS2.

[0039] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the invention belongs.

[0040] As used herein, “allele” refers to any of one or more alternative forms of a gene locus, all of which alleles relate to a trait or characteristic. In a diploid cell or organism, the two alleles of a given gene occupy corresponding loci on a pair of homologous chromosomes.

[0041] As used herein, “gene” refers to a nucleic acid sequence that encompasses a 5' promoter region associated with the expression of the gene product, any intron and exon regions and 3' or 5' untranslated regions associated with the expression of the gene product.

[0042] As used herein, “genotype” refers to the genetic constitution of a cell or organism.

[0043] The term "heterologous" or "exogenous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous or exogenous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

[0044] "Operable linkage" or "operably linked" as used herein is understood as meaning, for example, the sequential arrangement of a promoter and the nucleic acid to be expressed and, if appropriate, further regulatory elements such as, for example, a terminator, in such a way that each of the regulatory elements can fulfill its function in the recombinant expression of the nucleic acid to make RNA. This does not necessarily require direct linkage in the chemical sense. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions which are somewhat distant, or indeed from other DNA molecules (cis or trans localization). Preferred arrangements are those in which the nucleic acid sequence to be expressed recombinantly is positioned downstream of the sequence which acts as promoter, so that the two sequences are covalently bonded with one another.

[0045] As used herein, "phenotype" refers to the detectable characteristics of a cell or organism, which characteristics are the manifestation of gene expression.

[0046] The terms "polynucleotide," "nucleic acid" and "nucleic acid molecule" are used interchangeably herein to refer to a polymer of nucleotides which may be a natural or synthetic linear and sequential array of nucleotides and/or nucleosides, including deoxyribonucleic acid, ribonucleic acid, and derivatives thereof. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role. Unless otherwise indicated, nucleic acids or polynucleotide are written left to right in 5' to 3' orientation. Nucleotides are referred to by their commonly accepted single-letter codes. Numeric ranges are inclusive of the numbers defining the range. The "nucleic acid" may also optionally contain non-naturally occurring or altered nucleotide bases that permit correct read through by a polymerase and do not reduce expression of the nucleic acid.

[0047] The terms "polypeptide," "peptide," and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally

occurring amino acid, as well as to naturally occurring amino acid polymers. Amino acids may be referred to by their commonly known three-letter or one-letter symbols. Amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range.

[0048] As used herein, the term "substantially homologous" or "substantial homology", with reference to a nucleic acid sequence, includes a nucleotide sequence that hybridizes under stringent conditions to a referenced SEQ ID NO:, or a portion or complement thereof, are those that allow an antiparallel alignment to take place between the two sequences, and the two sequences are then able, under stringent conditions, to form hydrogen bonds with corresponding bases on the opposite strand to form a duplex molecule that is sufficiently stable under conditions of appropriate stringency, including high stringency, to be detectable using methods well known in the art. Substantially homologous sequences may have from about 70% to about 80% sequence identity, or more preferably from about 80% to about 85% sequence identity, or most preferable from about 90% to about 95% sequence identity, to about 99% sequence identity, to the referent nucleotide sequences as set forth the sequence listing, or the complements thereof. Alternatively, substantially homologous sequences include those which hybridize under stringent conditions to the target regions of introns of plant genes. As used herein, stringent conditions are those that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% NaDodSO₄ at 50° C.; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42° C.; or (3) employ 50% formamide, 5xSSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5xDenhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS and 10% dextran sulfate at 42° C. in 0.2xSSC and 0.1% SDS. For stringency conditions, see also U.S. Patent Nos. 8,455,716 and 8,536,403.

[0049] As used herein, the term "sequence identity", "sequence similarity" or "homology" is used to describe sequence relationships between two or more nucleotide sequences. The percentage of "sequence identity" between two sequences is determined by comparing two optimally aligned sequences over a comparison window such as the full length of a referenced SEQ ID NO:, wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is

calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. A sequence that is identical at every position in comparison to a reference sequence is said to be identical to the reference sequence and vice-versa. A first nucleotide sequence when observed in the 5' to 3' direction is said to be a "complement" of, or complementary to, a second or reference nucleotide sequence observed in the 3' to 5' direction if the first nucleotide sequence exhibits complete complementarity with the second or reference sequence. As used herein, nucleic acid sequence molecules are said to exhibit "complete complementarity" when every nucleotide of one of the sequences read 5' to 3' is complementary to every nucleotide of the other sequence when read 3' to 5'. A nucleotide sequence that is complementary to a reference nucleotide sequence will exhibit a sequence identical to the reverse complement sequence of the reference nucleotide sequence. These terms and descriptions are well defined in the art and are easily understood by those of ordinary skill in the art.

[0050] As used herein, a "comparison window" or "window of comparison" refers to a conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150, in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (i.e. gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Those skilled in the art should refer to the detailed methods used for sequence alignment, such as in the Wisconsin Genetics Software Package Release 7.0 (Genetics Computer Group, 575 Science Drive Madison, Wis., USA).

[0051] As described herein, the chemical composition of floral VOC was determined at 4 different stages of flower formation and performed RNA-seq on mature yellow flowers of *C. odorata* var. *fruticosa* where the production of floral VOCs is at the maximum. Terpenes formed the bulk of floral VOCs. The transcriptome data revealed 16 TPS transcripts from dwarf ylang ylang flowers out of which 4 were functionally characterized. Two TPSs were identified as mono-TPSs, CoTPS1 and CoTPS4, which catalyze the formation of multiproduct β -thujene/sabinene/ β -pinene/ α -terpinene and geraniol, respectively. The other two CoTPS2 and CoTPS3 were ascertained as sesqui-TPSs. CoTPS3 was shown to catalyze the formation of α -bergamotene. CoTPS2 was found to be a multifunctional and novel TPS which could synthesize

three sesquiterpene compounds, β -ylangene/ β -copaene/ β -cubebene from the farnesyl pyrophosphate (FPP) substrate.

[0052] In a first aspect, the present invention provides an isolated nucleic acid encoding a CoTPS protein. In one embodiment, the nucleic acid further comprises a plant operable promoter operably linked to the coding sequence. In some embodiments, the CoTPS protein is CoTPS2 protein comprising the amino acid sequence set forth in SEQ ID NO:10. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:9. In a further embodiment, the nucleic acid encodes a variant CoTPS2 protein. In one embodiment, the variant CoTPS2 protein has at least 90% sequence identity with the CoTPS2 protein while having the activity of the CoTPS2 protein. In another embodiment, the variant CoTPS2 protein has one or more amino acid changes in the amino acid sequence of the CoTPS2 protein while having the activity of the CoTPS2 protein.

[0053] In other embodiments, the CoTPS protein is CoTPS3 protein comprising the amino acid sequence set forth in SEQ ID NO:6. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:5. In a further embodiment, the nucleic acid encodes a variant CoTPS3 protein. In one embodiment, the variant CoTPS3 protein has at least 90% sequence identity with the CoTPS3 protein while having the activity of the CoTPS3 protein. In another embodiment, the variant CoTPS3 protein has one or more amino acid changes in the amino acid sequence of the CoTPS3 protein while having the activity of the CoTPS3 protein.

[0054] In some embodiments, the CoTPS protein is CoTPS1 protein comprising the amino acid sequence set forth in SEQ ID NO:8. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:7. In a further embodiment, the nucleic acid encodes a variant CoTPS1 protein. In one embodiment, the variant CoTPS1 protein has at least 90% sequence identity with the CoTPS1 protein while having the activity of the CoTPS1 protein. In another embodiment, the variant CoTPS1 protein has one or more amino acid changes in the amino acid sequence of the CoTPS1 protein while having the activity of the CoTPS1 protein.

[0055] In other embodiments, the CoTPS protein is CoTPS4 protein comprising the amino acid sequence set forth in SEQ ID NO:12. In one embodiment, the nucleic acid comprises the nucleotide sequence set forth in SEQ ID NO:11. In a further embodiment, the nucleic acid encodes a variant CoTPS4 protein. In one embodiment, the variant CoTPS4 protein has at least 90% sequence identity with the CoTPS4 protein while having the activity of the CoTPS4

protein. In another embodiment, the variant CoTPS4 protein has one or more amino acid changes in the amino acid sequence of the CoTPS4 protein while having the activity of the CoTPS4 protein.

[0056] In some embodiments, the polynucleotide may be one encoding a polypeptide of a variant of the amino acid sequences disclosed herein, which variant is an amino acid sequence disclosed herein having one or several amino acid residues substituted, deleted, inserted and/or added. The site at which one or several amino acid residues are substituted, deleted, inserted and/or added may be any site in the amino acid sequence, as long as the polypeptide with one or several amino acid residues substituted, deleted, inserted and/or added has the function of regulating the flowering time of a plant. As used herein, the term "one or several amino acid residues" refers specifically to up to 10 amino acid residues in number, preferably to up to 6 amino acid residues, more preferably to up to 2 amino acid residues and even more preferably to one amino acid residue.

[0057] When the amino acids are altered, for example, by substitution, it is preferable to be conservatively substituted. This means that a particular amino acid residue is substituted with a different amino acid in which the properties of the amino acid side-chain are conserved. Non-limited examples of such the conservative substitution include substitution between hydrophobic amino acids such as alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, tyrosine and valine, substitution between hydrophilic amino acids such as arginine, aspartic acid, asparagine, cysteine, glutamic acid, glutamine, glycine, histidine, lysine, serine and threonine, substitution between amino acids having an aliphatic side chain such as glycine, alanine, valine, leucine, isoleucine and proline, substitution between amino acids having a hydroxy-containing side chain such as serine, threonine and tyrosine, substitution between amino acids having a sulfur atom-containing side chain such as cysteine and methionine, substitution between amino acids having a carboxylic acid- and amide-containing side chain such as aspartic acid, asparagine, glutamic acid and glutamine, substitution between amino acids having a base-containing side chain such as arginine, lysine and histidine, and substitution between amino acids having an aromatic-containing side chain such as histidine, phenylalanine, tyrosine and tryptophan. The substitutions between amino acids having the same amino acid side-chain properties may retain the biological activity of the polypeptide.

[0058] In some embodiments, the polynucleotide may be a variant of a polynucleotide selected from the group consisting of the polynucleotides described herein, which variant has one to 30 nucleotides substituted, deleted, inserted and/or added. The site at which nucleotides

are substituted, deleted, inserted and/or added may be any site, as long as the polynucleotide with substituted, deleted, inserted and/or added nucleotides has the function of synthesizing the terpene or sesquiterpene described herein for each of the CoTPS proteins described herein.

[0059] Examples of methods for preparing a nucleic acid encoding a protein comprising altered amino acids are well known to those skilled in the art, including site-directed mutagenesis (Kramer and Fritz, 1987). Examples of specific methods for altering nucleotides also include methods using a commercially available kit (e.g. Transformer Site-Directed Mutagenesis Kit: Clontech; QuickChange Site Directed Mutagenesis Kit: Stratagene) and methods using polymerase chain reaction (PCR). These methods are well known to those skilled in the art. The amino acid sequence of a protein may also be mutated in nature due to the mutation of a nucleotide sequence. A nucleic acid encoding proteins having the amino acid sequence of a natural CoTPS protein (such as the CoTPS2, CoTPS3, CoTPS1 or CoTPS4 protein) wherein one or more amino acids are substituted, deleted, and/or added are also included in the polynucleotide of the present invention, so long as they encode a protein functionally equivalent to a natural CoTPS protein. Also natural CoTPS protein homologs in related ylang ylang plants which show high identities to the sequences of the CoTPS protein are also included in the polynucleotide of the present invention, so long as they encode a protein functionally equivalent to a natural CoTPS protein. Additionally, nucleotide sequence variants that do not give rise to amino acid sequence changes in the protein (degeneracy variants) are also included in the polynucleotide of the present invention.

[0060] In a second aspect, the present invention provides a construct or vector comprising an isolated nucleic acid as described herein. In one embodiment, the construct or vector is an expression construct or vector. In another embodiment, the construct or vector further comprises a selectable marker. In a further embodiment, the construct or vector comprises a recombination marker free system. In some embodiments, the recombination marker free system is a Cre-lox recombination marker free system, a Zinc finger marker free system, a TALE nucleases marker free system or a CRISPR-Cas marker free system.

[0061] The construct typically includes regulatory regions operatively linked to the 5' side of the nucleic acid described herein (such as a nucleic acid encoding a CoTPS2 protein, a nucleic acid encoding a CoTPS3 protein, a nucleic acid encoding a CoTPS1 protein or a nucleic acid encoding a CoTPS4 protein) and/or to the 3' side of the nucleic acid. A cassette containing all of these elements is also referred to herein as an expression cassette. The expression cassettes may additionally contain 5' leader sequences in the expression cassette construct. The

regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or the polynucleotide encoding a signal anchor may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or the polynucleotide encoding a signal anchor may be heterologous to the host cell or to each other. See, U.S. Patent No. 7,205,453 and U.S. Patent Application Publication Nos. 2006/0218670, 2006/0248616 and 20090100536, and the references cited therein. The expression cassettes may additionally contain 5' leader sequences in the expression cassette construct. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include those described in International Publication No. WO 2008/094127 and the references cited therein. In some embodiments, the nucleic acid construct further comprises a plant operable terminator. Plant operable terminators are well known in the art. In one embodiment, the terminator is the native *CoTPS* terminator.

[0062] A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. That is, the nucleic acids can be combined with constitutive, tissue-preferred, or other promoters for expression in the host cell of interest. Such constitutive promoters include, for example, the core promoter of the *Rsyn7* (WO 99/48338 and U.S. Pat. No. 6,072,050); the core *CaMV 35S* promoter (Odell et al., 1985); rice actin (McElroy et al., 1990); ubiquitin (Christensen and Quail, 1989; Christensen et al., 1992); *pEMU* (Last et al., 1991); *MAS* (Velten et al., 1984); *ALS* promoter (U.S. Pat. No. 5,659,026), and the like. Other constitutive promoters include, for example, those disclosed in U.S. Pat. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142.

[0063] Other promoters include inducible promoters, particularly from a pathogen-inducible promoter. Such promoters include those from pathogenesis-related proteins (PR proteins), which are induced following infection by a pathogen; e.g., PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, etc. Other promoters include those that are induced locally at or near the site of pathogen infection. In further embodiments, the promoter may be a wound-inducible promoter. In other embodiments, chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. The promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. In addition, tissue-preferred promoters can be utilized to target enhanced expression of a polynucleotide of interest within a particular plant tissue. Each of these

promoters are described in U.S. Pat. Nos. 6,506,962, 6,575,814, 6,972,349 and 7,301,069 and in U.S. Patent Application Publication Nos. 2007/0061917 and 2007/0143880.

[0064] Generally, the expression cassette may additionally comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Usually, the plant selectable marker gene will encode antibiotic resistance, with suitable genes including at least one set of genes coding for resistance to the antibiotic spectinomycin, the streptomycin phosphotransferase (spt) gene coding for streptomycin resistance, the neomycin phosphotransferase (nptII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (hpt or aphiv) gene encoding resistance to hygromycin, acetolactate synthase (als) genes. Alternatively, the plant selectable marker gene will encode herbicide resistance such as resistance to the sulfonylurea-type herbicides, glufosinate, glyphosate, ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D), including genes coding for resistance to herbicides which act to inhibit the action of glutamine synthase such as phosphinothricin or basta (e.g., the bar gene). See generally, International Publication No. WO 02/36782, U.S. Patent No. 7,205,453 and U.S. Patent Application Publication Nos. 2006/0218670, 2006/0248616, 2007/0143880 and 2009/0100536, and the references cited therein. See also, Jefferson et al. (1987); De Wet et al. (1987); Goff et al. (1990); Kain et al. (1995) and Chiu et al. (1996). This list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used. The selectable marker gene is also under control of a promoter operable in the plant species to be transformed. Such promoters include those described in International Publication No. WO 2008/094127 and the references cited therein.

[0065] Alternatively, the expression cassette may additionally comprise a recombination marker free system. In some embodiments, the recombination marker free system is a Cre-lox recombination marker free system, a Zinc finger marker free system, a TALE nucleases marker free system or a CRISPR-Cas marker free system. Such a system is useful for producing selection marker free transgenic oil palm plants or other plants.

[0066] In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro*

mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g. transitions and transversions may be involved.

[0067] In a third aspect, the present invention provides a transgenic plant comprising a nucleic acid, construct or vector described herein. In one embodiment, the transgenic plant may be any plant species. In another embodiment, the transgenic plant may be a member of the *Cananga* genus. In an additional embodiment, the transgenic plant is a *Cananga odorata* plant. In a further embodiment, the transgenic plant may be *Cananga odorata* var. fruticosa (dwarf ylang ylang). In some embodiments, the transgenic *Cananga* plant or *Cananga odorata* plant or *Cananga odorata* var. fruticosa plant overexpresses a CoTPS protein compared to a non-transgenic plant. In these embodiments, a nucleic acid, nucleic acid construct or expression vector as described above is introduced into a plant cell by conventional techniques such as those described below or those well known to the skilled artisan to produce transgenic plants that overexpress a CoTPS protein.

[0068] In other embodiments, the transgenic plant expresses a lower amount of a CoTPS protein compared to a non-transgenic plant, i.e., the expression of a *CoTPS* gene is down regulated. A reduced expression level of a CoTPS protein be brought about by using well known techniques, including, but not limited to, RNAi techniques, such as dsRNA, miRNA, siRNA, smRNA, hpRNA or ihpRNA (collectively referred to as RNAi molecules), sense suppression (co-suppression), antisense, and the like. Such techniques are described in U.S. Patent Nos. 7,312,323 and references cited therein. For example, reduction might be accomplished, for example, with transformation of a plant cell to comprise a promoter and other 5' and/or 3' regulatory regions described herein linked to an antisense nucleotide sequence, hairpin, RNA interfering molecule, double stranded RNA, microRNA or other nucleic acid molecule, such that tissue-preferred expression of the molecule interferes with translation of the mRNA of the native DNA sequence or otherwise inhibits expression of the native DNA sequence in plant cells. For further description of RNAi techniques or microRNA techniques, see, e.g., U.S. Patent Nos. 5,034,323; 6,326,527; 6,452,067; 6,573,099; 6,753,139; and 6,777,588. See also International Publication Nos. WO 97/01952, WO 98/36083, WO 98/53083, WO 99/32619 and WO 01/75164; and U.S. Patent Application Publication Nos. 2003/0175965, 2003/0175783, 2003/0180945, 2004/0214330, 2005/0244858, 2005/0277610, 2006/0130176, 2007/0265220, 2008/0313773, 2009/0094711, 2009/0215860, 2009/0308041, 2010/0058498 and 2011/0091975. RNAi molecules or microRNA molecules (referred to collectively herein as RNAi molecules) can be prepared by the skilled artisan using techniques well known in the art,

including techniques for the selection and testing of RNAi molecules and microRNA molecules that are useful for down regulating a *CoTPS* genes. See, for example, Wesley et al. (2001), Mysara et al. (2011) and Yan et al. (2012).

[0069] It has typically been found that dsRNA of 200-700 bp are particularly suited for inducing RNAi in plants. It has also been found that hairpin RNAs containing an intron, for example, a construct comprising an RNA encoding sequence in a sense direction operably linked to an intron operably linked to an RNA encoding sequence in an antisense direction or *vice versa* which is capable of forming an intron-hairpin RNA (ihpRNA), is suitable for inducing RNAi in plants. See, for example, Wang et al. (2000), Fuentes et al. (2006), Bonfim et al. (2007) Vanderschuren et al. (2007a, 2007b), Zrachya et al. (2007). For example, a nucleic acid construct can be prepared that includes a nucleic acid that is transcribed into an RNA that can anneal to itself, e.g., a double stranded RNA having a stem-loop structure. In addition, hairpin structures can be prepared as described by Guo et al. (2003).

[0070] For example, a nucleic acid construct can be prepared that includes a nucleic that is transcribed into an RNA that can anneal to itself, e.g., a double stranded RNA having a stem-loop structure. In some embodiments, one strand of the stem portion of a double stranded RNA comprises a sequence that is similar or identical to the sense coding sequence, or a fragment thereof, of a *CoTPS* as described herein, and that is from about 10 nucleotides to about 1,800 nucleotides in length. The length of the sequence that is similar or identical to the sense coding sequence can be from 10 nucleotides to 1000 nucleotides, from 15 nucleotides to 600 nucleotides, from 20 nucleotides to 500 nucleotides, or from 25 nucleotides to 100 nucleotides, or any length within the 10 nucleotides to 2,500 nucleotides. The other strand of the stem portion of a double stranded RNA comprises a sequence that is similar or identical to the antisense strand, or a fragment thereof, of the coding sequence of the polypeptide of interest, and can have a length that is shorter, the same as, or longer than the corresponding length of the sense sequence. In some cases, one strand of the stem portion of a double stranded RNA comprises a sequence that is similar or identical to the 3' or 5' untranslated region, or a fragment thereof, of the mRNA encoding the *CoTPS*, and the other strand of the stem portion of the double stranded RNA comprises a sequence that is similar or identical to the sequence that is complementary to the 3' or 5' untranslated region, respectively, or a fragment thereof, of the mRNA encoding the *CoTPS*. In other embodiments, one strand of the stem portion of a double stranded RNA comprises a sequence that is similar or identical to the sequence of an intron or a fragment thereof in the pre-mRNA encoding the *CoTPS*, and the other strand of the stem portion

comprises a sequence that is similar or identical to the sequence that is complementary to the sequence of the intron or fragment thereof in the pre-mRNA.

[0071] The loop portion of a double stranded RNA can be from 3 nucleotides to 5,000 nucleotides, e.g., from 3 nucleotides to 2500 nucleotides, from 15 nucleotides to 1,000 nucleotides, from 20 nucleotides to 500 nucleotides, or from 25 nucleotides to 200 nucleotides, or any length within the 3 nucleotides to 5,000 nucleotides. The loop portion of the RNA can include an intron or a fragment thereof. A double stranded RNA can have zero, one, two, three, four, five, six, seven, eight, nine, ten, or more stem-loop structures.

[0072] Once a nucleic acid has been cloned into an expression vector, it may be introduced into a plant cell using conventional transformation procedures. Similarly, once a nucleic acid construct has been prepared that contains or encodes a molecule which suppresses or inhibits the expression of a *CoTPS* gene, the nucleic acid may be introduced into a plant cell using conventional transformation procedures. The term "plant cell" is intended to encompass any cell derived from a plant including undifferentiated tissues such as callus and suspension cultures, as well as plant seeds, pollen or plant embryos. Plant tissues suitable for transformation include leaf tissues, root tissues, meristems, protoplasts, hypocotyls, cotyledons, scutellum, shoot apex, root, immature embryo, pollen, and anther. "Transformation" means the directed modification of the genome of a cell by the external application of recombinant DNA from another cell of different genotype, leading to its uptake and integration into the subject cell's genome. In this manner, genetically modified plants, plant cells, plant tissue, seed, and the like can be obtained.

[0073] DNA constructs containing the nucleic acids of the present invention encoding the CoTPS proteins can be used to transform any monocot or dicot plant and particularly ylang ylang varieties. The constructs may be introduced into the genome of the desired plant host by a variety of conventional techniques. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. Transformation protocols may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation, as is well known to the skilled artisan. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent

marker into the plant cell DNA when the cell is infected by the bacteria. Thus, any method, which provides for effective transformation/transfection may be employed. See, for example, U.S. Patent Nos. 7,241,937, 7,273,966 and 7,291,765 and U.S. Patent Application Publication Nos. 2007/0231905 and 2008/0010704 and references cited therein. See also, International Published Application Nos. WO 2005/103271 and WO 2008/094127 and references cited therein.

[0074] Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype, e.g., a transgenic plant. A "transgenic plant" is a plant into which foreign DNA has been introduced. A "transgenic plant" encompasses all descendants, hybrids, and crosses thereof, whether reproduced sexually or asexually, and which continue to harbor the foreign DNA. Regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. See for example, International Published Application No. WO 2008/094127 and references cited therein.

[0075] The foregoing methods for transformation are typically used for producing a transgenic variety in which the expression cassette is stably incorporated. After the expression cassette is stably incorporated in transgenic plants, it can be transferred to other plants by sexual crossing. In one embodiment, the transgenic variety could then be crossed, with another (non-transformed or transformed) variety, in order to produce a new transgenic variety. Alternatively, a genetic trait which has been engineered into a particular ylang ylang line using the foregoing transformation techniques could be moved into another line using traditional backcrossing techniques that are well known in the plant breeding arts. For example, a backcrossing approach could be used to move an engineered trait from a public, non-elite variety into an elite variety, or from a variety containing a foreign gene in its genome into a variety or varieties which do not contain that gene. As used herein, "crossing" can refer to a simple X by Y cross, or the process of backcrossing, depending on the context. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0076] Once transgenic plants of this type are produced, the plants themselves can be cultivated in accordance with conventional procedures. Transgenic seeds can, of course, be recovered from the transgenic plants. These seeds can then be planted in the soil and cultivated using conventional procedures to produce transgenic plants. The cultivated transgenic plants

will express the DNA of interest in a tissue-preferred or tissue-specific manner as described herein.

[0077] In a fourth aspect, the present invention provides a method for manipulating synthesis of a terpene or sesquiterpene in a transgenic plant to alter fragrance/flavor characteristics and/or plant pathogen interactions and/or insect interactions. In one embodiment, the transgenic plant may be any plant species. In another embodiment, the transgenic plant may be a *Cananga odorata* plant. In a further embodiment, the transgenic plant may be *Cananga odorata* var. *fruticosa* (dwarf ylang ylang). In one embodiment, the synthesis of the sesquiterpenes β -ylangene, β -copaene and β -cubebene is manipulated by overexpressing the CoTPS2 protein in the transgenic plant compared to a non-transgenic plant. In another embodiment, the synthesis of the sesquiterpene α -bergamotene is manipulated by overexpressing the CoTPS3 protein in the transgenic plant compared to a non-transgenic plant. In an additional embodiment, the synthesis of multiproduct β -thujene/sabinene/ β -pinene/ α -terpinene is manipulated by overexpressing the CoTPS1 protein in the transgenic plant compared to a non-transgenic plant. In a further embodiment, the synthesis of the terpene geraniol is manipulated by overexpressing the CoTPS4 protein in the transgenic plant compared to a non-transgenic plant. In some embodiments, the synthesis of the described terpenes or sesquiterpenes is manipulated by reducing expression of the corresponding CoTPS protein in the transgenic plant compared to a non-transgenic plant.

[0078] In one embodiment, an expression vector described herein is introduced into plant cells to obtain a transgenic plant, in which the synthesis of the terpene (β -thujene/sabinene/ β -pinene/ α -terpinene or geraniol) or sesquiterpene (α -bergamotene or β -ylangene/ β -copaene/ β -cubebene) is manipulated. Even a transgenic plant is not necessary to be obtained. Depending on the host plant, the nucleic acids of the present invention may be introduced into plant cells such that the nucleic acid encoding a flowering protein described herein can be expressed in the plant cells.

[0079] In another embodiment, a nucleotide sequence is integrated into the genomic DNA of a plant to enhance the expression of the endogenous *CoTPS* gene. Preferably, examples of such a nucleotide sequence include an expression control sequence. More specifically, examples of such an expression control sequence include promoter sequences and enhancer sequences. Such an expression control sequence is operably integrated into the genomic DNA of a plant to enhance the expression of the endogenous *CoTPS* gene in the plant. As described herein, enhanced expression of *CoTPS2* gene in the plant will enhance synthesis of β -ylangene, β -copaene and β -cubebene in the plant. Similarly, enhanced expression of *CoTPS3* gene in the

plant will enhance synthesis of α -bergamotene in the plant. Similarly, enhanced expression of *CoTPS1* gene in the plant will enhance synthesis of β -thujene/sabinene/ β -pinene/ α -terpinene in the plant. Similarly, enhanced expression of *CoTPS4* gene in the plant will enhance synthesis of geraniol in the plant.

[0080] In a fifth aspect, the present invention provides a method of preparing at least one terpene or sesquiterpene comprising the steps of (a) culturing a cell which has been genetically modified with a polynucleotide of the invention to provide CoTPS activity and (b) separating the at least one terpene or sesquiterpene produced. In one embodiment, the cells may be any cell type that can be grown in culture. In another embodiment, the cells may be bacteria or yeast cells for producing the terpene or sesquiterpene. In one embodiment, the at least one sesquiterpene is β -ylangene, β -copaene and/or β -cubebene and the CoTPS activity is CoTPS2 activity. In another embodiment, the at least one sesquiterpene is α -bergamotene and the CoTPS activity is CoTPS3 activity. In an additional embodiment, the at least one terpene is β -thujene/sabinene/ β -pinene/ α -terpinene and the CoTPS activity is CoTPS1 activity. In a further embodiment, the at least one terpene is geraniol and the CoTPS activity is CoTPS4 activity. In one embodiment, the cell contains a metabolic pathway for producing farnesyl pyrophosphate which provides the farnesyl pyrophosphate that is the substrate for CoTPS2 or CoTPS3. In another embodiment, the cell has been transformed with a nucleic acid encoding farnesyl pyrophosphate synthase to produce farnesyl pyrophosphate which provides the farnesyl pyrophosphate that is the substrate for CoTPS2 or CoTPS3. In one embodiment, the cell contains a metabolic pathway for producing geranyl diphosphate which provides the geranyl diphosphate that is the substrate for CoTPS1 or CoTPS4. In another embodiment, the cell has been transformed with a nucleic acid encoding geranyl diphosphate synthase to produce geranyl diphosphate which provides the geranyl diphosphate that is the substrate for CoTPS1 or CoTPS4. Techniques for synthesizing terpenes and/or sesquiterpenes in cell culture are described in, for example U.S. Patent Nos. 7,453,024 and 8,062,878, U.S. Patent Application Publication Nos. 2012/0107893 and 2013/0302861 and International Published Application No. WO 2011/074954.

[0081] In this aspect of the invention, the nucleic acid construct contains a promoter that is operably in the host cell and that is operatively linked to the CoTPS2 encoding nucleic acid described herein. Examples of suitable yeast promoters include, but not limited to, copper inducible promoter CUP1 (P_{CUP1}) as well as promoters for pyruvate decarboxylase (PDC1), phosphoglycerate kinase (PGK), xylose reductase (XR), xylitol dehydrogenase (XDH), L-(+)-

lactate-cytochrome c oxidoreductase (CYB2), translation elongation factor-1 (TEF1) and translation elongation factor-2 (TEF2) genes. Additional yeast promoters include the GAP promoter, GAL1 promoter, AOX1 promoter, FLD1 promoter, ADH1 promoter, GAL3 promoter, GAL4 promoter, GAL7 promoter, CTR1 promoter, CTR3 promoter, MET3 promoter and TDH1 promoter. The nucleic acid construct may also contain a terminator. Examples of yeast terminators include, but not limited to, terminators for ADH1, TDH1, pyruvate decarboxylase (PDC1), xylose reductase, (XR), xylitol dehydrogenase (XDH), L-lactate:ferricytochrome c oxidoreductase (CYB2) or iso-2-cytochrome c (CYC) genes (e.g., terminator of CYC1 (T_{CYC1})), or a terminator from the galactose family of genes in yeast, particularly the GAL10 terminator and GAL80 terminator.

[0082] Various methods can be used to introduce the expression vector of some embodiments of the invention into bacterial or yeast cells. Such methods are generally described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Laboratory, New York (1982, 1989, 2001, 2012), in Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1989), Chang et al., *Somatic Gene Therapy*, CRC Press, Ann Arbor, Mich. (1995), Vega et al., *Gene Targeting*, CRC Press, Ann Arbor Mich. (1995), *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, Butterworths, Boston Mass. (1988) and Gilboa et al. (*Biotechniques* 4 (6): 504-512, 1986) and include, for example, stable or transient transfection, lipofection, electroporation and infection with recombinant viral vectors. In addition, see U.S. Patent Nos. 5,464,764 and 5,487,992 for positive-negative selection methods.

[0083] The method further comprises providing and/or maintaining conditions suitable for sesquiterpene production within the yeast cell, e.g. cultivating the yeast under conditions conducive to the production of the sesquiterpenes, prior to isolating the sesquiterpenes. These conditions are known to the skilled person. Generally, they may be adjusted by selection of an adequate medium, temperature, and pH.

[0084] The sesquiterpenes may be isolated from the culture medium or the yeast cells by any method used in the art including, but not limited to, chromatography, extraction, in-situ product removal and distillation.

[0085] It will also be appreciated that transformation of other non-plant hosts is feasible, including well known prokaryotic and eukaryotic cells, such as bacteria (e.g. *E. coli*, *Agrobacterium*, *Lactobacillus*), fungi, yeast, insect cells and animal cells. Such transformation enables production of recombinant polypeptides of the invention. Cell free systems (e.g. Roche Rapid Translation System) for production of recombinant proteins can also be used (Zubay

Annu Rev Genet 7, 267-287 (1973)). The polypeptides of the invention produced in any such hosts may be isolated and purified from same using well known techniques. The polypeptides may be used in cell-free systems for synthesis of sesquiterpenes for flavoring or scent uses, or for use in pheromone or antimicrobial agents. Such compounds could be β -ylangene, β -copaene or β -cubebene.

[0086] The pleasant fragrance of ylang ylang varieties (*Cananga odorata*) are mainly due to volatile organic compounds (VOCs) produced by the flowers. Floral scents are a key factor in plant-insect interactions and are vital for successful pollination. *C. odorata* var. *fruticosa* or dwarf ylang ylang is a variety of ylang ylang which is popularly grown in Southeast Asia as a small and compact shrub with aromatic flowers. The following Examples describe the combined use of bioinformatics and chemical analysis to discover the VOC biosynthesis pathways and related genes. The scented flowers of *C. odorata* var. *fruticosa* were analysed by GC-MS and a total of 49 volatile organic compounds were identified at four different stages of flower development. The bulk of these VOCs were terpenes - mainly sesquiterpenes. The various terpene synthases (TPSs) involved in the production of these essential oils was identified using RNA-seq on mature flowers. From the RNA-seq data, 4 full-length TPSs were functionally characterized. *In vitro* assays showed that recombinant proteins from two TPSs, CoTPS1 and CoTPS4 synthesized β -thujene/sabinene/ β -pinene/ α -terpinene and geraniol from geranyl diphosphate, respectively. Accordingly, they were classified as mono-TPSs. The other two TPSs were identified as sesqui-TPSs. CoTPS3 catalysed the conversion of farnesyl diphosphate to α -bergamotene, whereas CoTPS2 was found to be a multifunctional and novel TPS which could catalyse the synthesis of three sesquiterpene compounds, β -ylangene/ β -copaene/ β -cubebene. Additionally, *in vitro* results of the two sesqui-TPSs were confirmed *in planta* by transient expression of these TPSs in *N. benthamiana* leaves by *Agrobacterium*-mediated infiltration.

[0087] The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, immunology, cell biology, cell culture and transgenic biology, which are within the skill of the art. See, e.g., Maniatis *et al.*, 1982, *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York); Sambrook *et al.*, 1989, *Molecular Cloning*, 2nd Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York); Sambrook and Russell, 2001, *Molecular Cloning*, 3rd Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York); Green and Sambrook, 2012, *Molecular Cloning*, 4th Ed. (Cold Spring

Harbor Laboratory Press, Cold Spring Harbor, New York); Ausubel *et al.*, 1992), *Current Protocols in Molecular Biology* (John Wiley & Sons, including periodic updates); Glover, 1985, *DNA Cloning* (IRL Press, Oxford); Russell, 1984, *Molecular biology of plants: a laboratory course manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); Guthrie and Fink, *Guide to Yeast Genetics and Molecular Biology* (Academic Press, New York, 1991); Harlow and Lane, 1988, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York); *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Riott, *Essential Immunology*, 6th Edition, Blackwell Scientific Publications, Oxford, 1988; Fire et al., *RNA Interference Technology: From Basic Science to Drug Development*, Cambridge University Press, Cambridge, 2005; Schepers, *RNA Interference in Practice*, Wiley-VCH, 2005; Engelke, *RNA Interference (RNAi): The Nuts & Bolts of siRNA Technology*, DNA Press, 2003; Gott, *RNA Interference, Editing, and Modification: Methods and Protocols (Methods in Molecular Biology)*, Human Press, Totowa, NJ, 2004; Sohail, *Gene Silencing by RNA Interference: Technology and Application*, CRC, 2004.

EXAMPLES

[0088] The present invention is described by reference to the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below were utilized

EXAMPLE 1

Materials and Methods

[0089] *Plant materials:* *Cananga odorata* var. *fruticosa* (known as dwarf ylang ylang) was grown in Singapore were collected in November for RNA-seq. Four different stages of flowers and leaves of dwarf ylang ylang were obtained in April for further experiments. The four

different stages of flowers are as follows: (B) bud stage; completely closed petal, green; (I) initial-flowering stage: semi-open small and short petals, green, 9 days after bud stage; (II) full-flowering stage: completely open large and long petals, yellowish green, 20 days after bud stage; (III) end-flowering stage: fully matured petals, yellow, 30 days after bud stage.

Four-week old *Nicotiana benthamiana* plants grown in a greenhouse were used for *in vivo* characterization and subcellular localization of CoTPSs.

[0090] *Essential oils extraction from ylang ylang flowers:* Flowers and leaves from ylang ylang were frozen in liquid nitrogen and ground to a powder by pre-chilled mortar and pestle. About 500 mg of powder was dissolved in 500 μ l ethyl acetate (Fisher Scientific) including 1 μ l (10 mg/ml) of camphor (Sigma-Aldrich) as internal standard. The slush was vortexed and incubated on a horizontal shaker at 50 rpm for 2 hr. After centrifugation of the mixture at 13,000 g for 10 min, the resulting ethyl acetate upper layer extract was transferred into a new Eppendorf tube and mixed with 300 mg anhydrous Na₂SO₄ (Sigma-Aldrich) to remove water. Following the second centrifugation, the extract was transferred into a 2 ml glass vial for gas chromatography-mass spectrometry (GC-MS) analysis (Agilent Technologies).

[0091] *RNA isolation for RNA sequencing:* Frozen ylang ylang flowers were homogenized using pre-chilled mortar and pestle into a fine powder, and total RNA was isolated using the TRIzol method (Invitrogen). Purified RNA samples were first treated with RNase-free DNaseI (Roche) to remove genomic DNA and then extracted by chloroform. The RNA quantity was determined with Nanodrop spectrophotometer, ND-1000 (Thermo Fisher Scientific). The RNA integrity number (RIN) was evaluated using Agilent 2100 bioanalyzer and RNA 6000 Nano Labchip Kit (Agilent Technologies). RNA with a RIN value of $7 < x < 10$ was sent for RNA-seq to the Rockefeller University Genomics Resource Center (New York, USA). RNA-seq and assembly were described in Jin et al (2014).

[0092] *Quantitative real time PCR (qRT-PCR):* Quantitative RT-PCR was performed to investigate gene expression pattern during ylang ylang flower development. One μ g of total RNA was used for first-strand cDNA synthesis with M-MLV reverse transcriptase (RT) (Promega). The qRT-PCR reactions were performed using Applied Biosystems (AB) 7900HT fast real-time PCR system and AB power SYBR green PCR master mix (Life technologies). Oligonucleotide primers for qRT-PCR were designed for selected genes of MEP, MVA, shikimate, benzyl and phenyl propanoid pathways using the Primer3 (<http://colonslashslashbioinfo.utdallas.edu/primer3-0.4.0/>) and listed in Table 1. Each PCR product obtained from regular PCR was cloned into pGEM-T easy vector (Promega) and then verified by

sequencing. Specificity of the amplified PCR product was assessed by a melting curve analysis. All experiments were carried out in technical triplicates and with biological duplicates. Non-template control was included for each gene to exclude random/reagent contamination and primer-dimer formation. A mock reaction containing all the RT-PCR reagents except the reverse transcriptase was used as a negative control. *Actin* gene was used as an internal normalization in each qRT-PCR.

TABLE 1
Oligonucleotide Primers

Name	Forward (SEQ ID NO:)	Reverse (SEQ ID NO:)
For qRT-PCR		
DXS1	TTTGATTGGGGTGAGAGAG (13)	GCCATCAAGGGACATGAACT (14)
DXS2	GGGGATCACAAAATCCATTG (15)	TGGGTAGCCTTTTCCCTTTT (16)
DXS3	CTTCTGCAAGCCCTAGATG (17)	GTTGCAGCAATGTGCCCTAGA (18)
DXS4	AAGGGAGAAATCCTTGCACT (19)	TGTTCTTGCAAAAGCTTCCT (20)
DXR	GTTCTGAGTGCAGCCCAATGA (21)	CAAGAACAGGTCGCAGATCA (22)
CMS	AGGAGCGTGTCACTGTTCT (23)	TGGCAAAATTCAGGTCAACA (24)
CMK	TCGGTTCTGATTGTCCCTTC (25)	CGAAGGCGCTTGTAACCTTC (26)
MCS	ACGCAAGCAAAAACCTCTCGT (27)	CGGTAGCGTTATCAGCATCA (28)
HDS	GTGGCTCCATTGAGGAAAA (29)	AATCCACAGTGGCAAGATCC (30)
ID11	ACTGTACAAGGCCCCCATCTG (31)	CCCACCAATTCAGAAAGGAAA (32)
ID12	CATGAGTTGCTCCTTCAGCA (33)	AAGCTTCCCTTTGTGCAGCAT (34)
AACT	TGCAGTGAAGTCTGCTGGTC (35)	CAACAGAGTTGCCACACAGC (36)
HMGs	CAAGACTTGGCTGATGCAAA (37)	AACAACAAGCCCAAAATCGTC (38)
HMGR	CAGGCTGATTCCCTGACAT (39)	CTGGCATGTGCATTGTATCC(40)
MVK	GGTGTTTCGGAGAGAGCTTG (41)	ACCCCCATGCACTGAAAGTAG (42)
PMK	TGCTCCTCGATGCTACAAATG (43)	AGGGTCTTCTCTGACCAGCA (44)

MPDC	CCCATCAATGACAGCATCAG (45)	CGTATTGGGATCTCCCTCAA (46)
DAHPS	ATCCTCCATCTTCCCTGCTT (47)	TTCTTGGATTTCAGCCATC (48)
DHQS	CCTTTCCTCCGCTTCTCT (49)	TTCCATGAACATGCCTTGA (50)
SDH	TGATCAGCTAGAGCGCAGAA (51)	CCATCTGTTCGACGGATTTT (52)
SK	CACCGTTCGAAAGATTGTT (53)	CCAGACCCCATCATACCAAC (54)
CS	GCTGCTGGAAGTTCTTTTGG (55)	TCTGTCTTGGCTCCTTCTA (56)
PAL	AATCTCTCAGGGGAAAGGAA (57)	CAACCCAGCGAGTTTACAT (58)
SAMT	TTAACGGCAGGAAGAGCCTA (59)	ATCACGGCCTTCATTTCTTG (60)
CHS	TTCTGTTCCTCAAGCGGAGT (61)	CGGAAGTAGCCACACATTCAT (62)
CHI	CATGAGCATGTCAGGAAGA (63)	GCACAGCAACTCATTTCCTCA (64)
F3H	TATCTCTGGTGGGATCGAC (65)	CTGCTTCTCCCTGAAAGATGG (66)
DFR	GCAATGGTGATTGGTCCTCT (67)	GAAAGCGGAATTGATGGGTAA (68)
C4H	ACCCACAAGCTTCCCTACCT (69)	AGGTACCAGGCATTACCCAG (70)
4CL	GTTGCACCTGCAGAACTTGA (71)	ACCTTCCGCGAGTCTCTTGAA (72)
CCR	GGAGCGATCTGGACTACTGC (73)	GAACATCGACGTAGGCCCTGT (74)
CVOMT	GCATGCTCACTCACTCTGGA (75)	CCCCAAAGACTTTTCCGTGA (76)
C3H	ACGCTCCAGGAGAAAGTACGA (77)	CACACGGTCCAATTCCCTCTT (78)
COMT	ACTCTCCAGTGCTGTGCTT (79)	AGCTGCCCGTTGTCTAGTGT (80)
CCOMT	TTGGATACCAGCGTCTACCC (81)	CTCACGGTTGATGTCCATTG (82)
CAD	AAAGTGGGAACAACGAATGC(83)	ACCGACCTCTTTGACCACAC (84)

TPS1	AGGGCCTTCCAGATTACATGAAGC (85)	AGCGTTGGAGTGTACTTGCTATTG (86)
TPS2	CATGATCATGGATGACACATATGATG (87)	CTCAACAGCAGTCAGAAAGGGCAC (88)
TPS3	AGGCTGTGGCTGAACCTCTA (89)	AGCAGATCCTTCTTGCTGGA (90)
TPS4	TAGCCATCTTACTGTGTGATTGACG (91)	GTAGTGTGTGACAAATGCCATGTAAAC (92)
actin	CTGGACGTGACCCTCACAGATGCT (93)	TCTTCTCAACAGAGGAGCTGCTCT (94)
For TOPO Cloning		
TPS1	CACCAATGGCCTTGAATACGTTCTTGCAATTTTC (95)	CTCAAGCACGATGGGTTC AACGGTAA GTGA (96)
TPS2	CACCAATGGCACTTATATTTGCAAAATGGCCACTCTGA (97)	ATCAACAAAAACGGTCCACTAGCACCATCGCA (98)
TPS3	CACCAATGAATCCTGTTTCTCTTTTGAGCTTATCAGGAG (99)	AAC TTGTTGGGAAAAATAGAGCCATGGCTTGA (100)
TPS4	CACCAATGGCTGCTACGAGAAACCTTCTTTACTT (101)	ATTGAAAGGCGTGAAAAACAAAGCTTGCA (102)
For <i>E. coli</i> expression vectors		
TPS1	aaGGATCCgATGGCCTTGAATACGTTCTTGCAATTTTC (103)	aaGGGGCCGCTTACTCAAGCACGATGGGTTC AAC (104)
TPS2	aaGTCGACATGGCACTTATATTTGCAAAATGGCCACTCTGA (105)	aaaGGGGCCGCTTAAATCAACAAAAACGGTCCACTAGCA (106)
TPS3	aaGGATCCgATGAATCCTGTTTCTCTTTTGAGCTTATCAGGAG (107)	aacGTCGACAACTTGTGTTGGGAAAAATAGAGCCATGGCT (108)
TPS4	aaaGGATCCtATGGCTGCTACGAGAAACCTTCTTTACTT (109)	aaaGTCGACATTGAAAGGCGTGAAAAACAAAGCTTGCA (110)

GGATCC, *Bam* HI; GCGGCCGC, *Not* I; GTCGAC, *Sal* I

[0093] *Sequence identification, multiple sequence alignments and phylogenetic analysis:* DNA sequences were edited and assembled using DNASTAR Lasergene 8 (DNASTAR, Inc.). The phylogenetic analysis of CoTPSs was performed using the maximum likelihood method in MEGA version 6 program (Tamura et al., 2011).

[0094] *Isolation of full-length cDNA of CoTPS and vector construction for Agrobacterium-mediated gene expression:* Full-length open reading frames (ORFs) of *CoTPSs* were amplified by PCR from dwarf ylang ylang flowers-derived cDNA with the primer sets listed in Table 1. Purified PCR products were cloned into pENTR/D-TOPO (Invitrogen). For YFP-fusion construct, the pTOPO clone harbouring each *CoTPS* gene was integrated into the destination vector, pBA-DC-YFP expression vector (Zhang et al., 2005), which contains the CaMV 35S promoter and C-terminal in frame *YFP*, to create *CoTPS-YFP* by LR Clonase (Invitrogen). All constructs were verified by DNA sequencing. The final plasmid was then transformed into GV3101 *Agrobacterium tumefaciens* by electroporation system (Bio-Rad), plated on LB plate containing spectinomycin (100 µg/ml) and gentamycin (20 µg/ml), and incubated at 28 °C for 2 days.

[0095] *Subcellular localization of CoTPSs:* To determine the subcellular localization of CoTPSs, *Agrobacterium*-mediated transient gene expression was performed using leaves of 4 week old *N. benthamiana* plant as described (Jin et al., 2014). Infiltrated *N. benthamiana* leaves expressing YFP-fused protein were excised, mounted onto slides, and imaged by a confocal laser scanning microscopy (Carl Zeiss LSM5 Exciter) with a standard filter set. Images were processed with LSM Image Browser (Carl Zeiss).

[0096] *Preparation of recombinant proteins:* To construct the vectors for the recombinant N-terminal poly-histidine (His)-tagged proteins, the full-length cDNAs of *CoTPS* was amplified by PCR with the primers designed restriction enzymes sites to the ends (Table 1). The PCR-amplified product and pET28b plasmid (Novagen) were separately digested with the corresponding restriction enzymes (New England Biolabs) and purified with Qiagen PCR purification kit (Qiagen). Then, the digested PCR product was cloned into pET28b expression vector using Rapid DNA ligation kit (Roche). The final construct was transformed into *E. coli* BL21(DE3)pLysS (Invitrogen) and recombinant proteins purified from *E. coli* extracts after isopropyl β-D-1-thiogalactopyranoside induction as described (Jang et al., 2005).

[0097] *In vitro TPS assay:* *In vitro* enzyme assay for TPS activity was performed in a final volume of 500 µl of reaction buffer (25 mM HEPES pH 7.4, 100 mM KCl, 7.5 mM MgCl₂, 5% (v/v) glycerol, 5 mM DTT), with about 20 µg of recombinant protein, and 10 µg of either

farnesyl pyrophosphate (FPP) or geranyl pyrophosphate (GPP) (Sigma-Aldrich). The reaction mixtures were gently mixed and carefully overlaid with 250 μ l hexane (Sigma-Aldrich) to trap volatile products. The tube was then sealed with parafilm and incubated at 30 °C for 2 hr, followed by 1 min vortex. After centrifugation at 1,200 g, 4 °C for 30 min, the hexane upper layer was transferred into a 2 ml glass vial for GC-MS analysis (see below *GC-MS analysis*). As a negative control, heat-inactivated recombinant protein was added to the enzyme assay.

[0098] *In vivo characterization of CoTPSs:* For *in vivo* TPS activity assay, *Agrobacterium* strain harbouring *CoTPS* construct was infiltrated with or without the strain carrying Arabidopsis HMGR (3-hydroxy-3-methylglutaryl-CoA reductase; *AtHMGR*) under the control of the *CaMV 35S* promoter into the underside of *N. benthamiana* leaves (Jin et al., 2014). All experiments were carried out with the viral-encoded protein P19 to improve transgene expression by suppressing post-transcriptional gene silencing (Voinnet et al, 2003). After infiltration, tobacco plants were maintained in a growth chamber at 25 °C, under long day condition (16 hr light/ 8 hr dark) for 3 days. Four to five infiltrated leaves were frozen immediately in liquid nitrogen and then homogenized with a pre-chilled mortar and pestle. Up to 400-600 mg of leaf powder were obtained from 4-5 leaves. Subsequent sample processing for GC-MS analysis was performed as described above (See *Essential oils extraction from ylang ylang flowers*). A *CaMV 35S::AtHMGR* construct served as a negative control.

[0099] *GC-MS analysis:* GC-MS analysis was performed on Agilent 7890A GC (Agilent Technologies) system and an Agilent-Technologies 5975C inert XL Mass selective detector, equipped with a HP-5MS UI column (30 m $\times$ 0.25 mm $\times$ 0.25 μ m; Agilent Technologies). Conditions were as follows: 5 μ l sample injection, splitless injection, oven program 50° C (1 min hold) at 8° C min⁻¹ to 300° C (5 min hold). For data processing, MSD ChemStation Data Analysis (Agilent Technologies) was used. The essential oil components were identified by comparison of their mass spectra with those in NIST 2011 library data of GC-MS system and comparison of their retention indices (RI). The retention indices (RI) were determined on the basis of n-alkanes (C8-C40) mix standard (Sigma-Aldrich) under the same operation conditions. Camphor was added to serve as an internal standard. The amount of each compound was calculated by measuring its peak area related to that of a known amount of camphor. The identified components along with their RI and relative % values are listed in Table 2.

EXAMPLE 2

Stage Specific Variations of Volatile
Organic Compounds in Dwarf Ylang Ylang Flowers

[0100] Flowers can emit different volatile compounds at different stages of development (Dudareva et al., 2000). The dwarf ylang ylang flowers have little floral scent when the petals are green but their scent gradually becomes stronger as the flower matures. To examine the overall intensity and the diversity of the floral scent during flower development, total essential oils from flowers at four different stages of development were analysed by gas chromatography-mass spectrometry (GC-MS) (Figures 1A and 1B). Figure 1B and Figure 9 show that the chemical composition of the essential oils from the floral bud stage to the 3 different stages of open flower development was very diverse both quantitatively and qualitatively. Only 15 compounds that have meaningful levels >0.1% of total volatile compounds were detected from floral buds (Figure 1B and Table 2). The number of peaks increased progressively during maturation of flower buds into fully open flowers. More than 20, 27, and 45 volatile compounds were obtained from the three different stages of flower development, undeveloped small flower (I), mature green flower (II) and fully-mature yellow flower (III). Hence, the fully-mature stage represents the stage where there was maximum production of VOCs by the flowers. At this stage majority of the volatiles were terpenes with a few benzenoids/phenolpropanoids compounds. Out of 45 compounds identified, 31 were identified as mono- and sesquiterpenes using the mass spectra reference library (Figure 1B and Table 2). Interestingly, over 90% of the total identified terpenes were sesquiterpenes consisting of α -farnesene (31.50%), α -bergamotene (26.79%), germacrene D (13.26%), β -caryophyllene (11.57%), humulene (1.63%), farnesol (0.75%), *trans*- β -farnesene (0.77%) and β -ylangen (0.63%), whereas monoterpenes were quantitatively less than 10%, mainly *cis*- β -ocimene (1.99%), *trans*- β -ocimene (1.55%) and β -linalool (1.40%). Other aromatic compounds were less than 3%. The relative amounts of all identified volatiles are shown in Table 2 (see mature yellow flowers, III). Dried yellow flowers of ylang ylang also showed similar volatile composition (Figure 10A). Compared to flowers, ylang ylang leaves contained very low levels of terpenes mainly comprising of α -pinene, β -caryophyllene, germacrene D, and phytol (Figure 10B).

TABLE 2
Essential Oils Composition of the
Flowers from *Cananga odorata* var. *fruticosa*

No ^a	Compounds	RT (min) ^b	RI ^c	Formula	Relative contents (%) ^d			
					Bud	I	II	III
1	β -Thujene	9.507	904	C ₁₀ H ₁₆	-	-	0.21	0.36
2	α -Pinene	9.652	919	C ₁₀ H ₁₆	3.08	1.30	0.99	0.65
3	Camphene	9.894	943	C ₁₀ H ₁₆	-	-	-	0.13
4	Sabinene	10.199	974	C ₁₀ H ₁₆	-	-	0.26	0.43
5	β -Pinene	10.342	988	C ₁₀ H ₁₆	-	-	0.45	0.66
6	α -Phellandrene	10.661	993	C ₁₀ H ₁₆	-	-	-	0.05
7	α -Terpinene	10.846	1012	C ₁₀ H ₁₆	-	-	-	0.27
8	p-Cresol methyl ether	10.884	1016	C ₈ H ₁₀ O	-	-	-	0.49
9	<i>trans</i> - β -Ocimene	11.054	1023	C ₁₀ H ₁₆	-	0.46	0.88	1.55
10	β -Ocimene	11.170	1024	C ₁₀ H ₁₆	-	0.40	1.09	1.99
11	γ -Terpinene	11.397	1027	C ₁₀ H ₁₆	-	-	-	0.15
12	Terpinolene	12.011	1032	C ₁₀ H ₁₆	-	-	-	0.16
13	β -Linalool	12.102	1040	C ₁₀ H ₁₈ O	-	0.38	0.99	1.40
14	Neo-allo-ocimene	12.594	1089	C ₁₀ H ₁₆	-	-	-	0.18
15	3,4-Dimethoxytoluene	14.460	1197	C ₉ H ₁₂ O ₂	-	-	-	0.39
16	2-Methoxy-4-vinylphenol	15.854	1276	C ₉ H ₁₀ O ₂	-	-	-	0.19
17	γ -Elemene	16.176	1306	C ₁₅ H ₂₄	-	-	-	0.16
18	Eugenol	16.481	1309	C ₁₀ H ₁₂ O ₂	-	-	-	0.17
19	α -Copaene	16.831	1321	C ₁₅ H ₂₄	1.81	1.24	0.83	0.44
20	γ -Gurjunene	17.025	1340	C ₁₅ H ₂₄	3.00	1.16	0.56	0.32
21	Methyleugenol	17.062	1344	C ₁₁ H ₁₄ O ₂	-	-	-	0.08
22	β -Caryophyllene	17.600	1398	C ₁₅ H ₂₄	16.47	23.50	15.90	11.57
23	β -Ylangene	17.683	1406	C ₁₅ H ₂₄	3.55	1.80	1.16	0.63
24	(E)- β -Farnesene	17.836	1421	C ₁₅ H ₂₄	-	0.51	0.61	0.77
25	γ -Muurolene	17.910	1429	C ₁₅ H ₂₄	1.21	0.60	0.38	0.20
26	Humulene	18.020	1440	C ₁₅ H ₂₄	3.44	3.44	2.24	1.63
27	β -Cubebene	18.110	1449	C ₁₅ H ₂₄	1.02	0.35	0.19	0.08
28	GermacreneD	18.580	1450	C ₁₅ H ₂₄	50.17	33.17	27.33	13.26
29	α -Farnesene	18.686	1461	C ₁₅ H ₂₄	-	10.03	19.89	31.50
30	α -Bergamotene	18.771	1469	C ₁₅ H ₂₄	-	18.67	23.63	26.79
31	Cedrene	18.932	1485	C ₁₅ H ₂₄	-	-	0.32	0.33
32	δ -Cadinene	18.989	1491	C ₁₅ H ₂₄	1.54	0.43	0.17	0.17
33	α -Patchoulene	19.035	1495	C ₁₅ H ₂₄	-	-	0.07	0.13
34	Elemicin	19.233	1515	C ₁₂ H ₁₆ O ₃	-	-	-	0.13
35	Germacrene D-4-ol	19.845	1537	C ₁₅ H ₂₆ O	0.75	0.21	0.16	0.13
36	β -Caryophyllene oxide	19.950	1548	C ₁₅ H ₂₄ O	-	-	-	0.07
37	Isoelemicin	20.561	1609	C ₁₂ H ₁₆ O ₃	-	-	0.15	0.13
38	Farnesol	21.527	1763	C ₁₅ H ₂₆ O	-	0.88	0.85	0.75
39	Benzyl benzoate	22.236	1812	C ₁₄ H ₁₂ O ₂	-	-	0.27	0.69
40	<i>cis</i> -11-Hexadecenal	22.505	1830	C ₁₆ H ₃₀ O	-	-	-	0.11
41	Octadecanal	22.623	1841	C ₁₈ H ₃₆ O	-	-	-	0.07
42	(E,E)-farnesol acetate	22.941	1873	C ₁₇ H ₂₈ O ₂	-	-	0.10	0.10
43	Z-9-Hexadecen-1-ol	23.328	1912	C ₁₆ H ₃₂ O	-	-	-	0.17
44	9-Nonadecene	23.475	1918	C ₁₉ H ₃₈	-	-	-	0.18
45	Benzyl salicylate	23.600	1931	C ₁₄ H ₁₂ O ₃	-	-	-	0.21

46	δ -Elemene	16.173	1308	$C_{15}H_{24}$	4.81	0.98	0.34	-
47	α -Ylangene	18.662	1458	$C_{15}H_{24}$	7.16	-	-	-
48	γ -Cadinene	18.910	1483	$C_{15}H_{24}$	1.31	0.52	-	-
49	α -Copaene-11-ol	21.094	1720	$C_{15}H_{24}O$	0.67	-	-	-

^aCompound listed in order of elution in a HP-5MS UI column; ^bRT, retention time (min); ^cRI, retention indices calculated against C_8 - C_{40} *n*-alkanes on the HP-5MS UI column; ^dRelative contents were done by calculating percentage of moles using internal standard, camphor (10 ug/ul).

[0101] Two sesquiterpenes, β -caryophyllene and germacrene D could be found at all stages and their levels were retained or slightly decreased during open flower development (#22, 28, Figure 1B and Figure 11). However, the other two major sesquiterpenes, α -farnesene and α -bergamotene were undetectable at the floral bud stage, but they were found at early flower development (#29, 30 in I, Figure 1B and Figure 11) and subsequently became the most abundant components of essential oils at mature stages of flower development (II and III, Figure 1B). Interestingly, most of the monoterpenes except α -pinene were undetectable at the floral bud stage, but they gradually increased during open flower development (Figure 1B). This increase in VOCs as the flower matures suggests the VOCs play a role in pollinator attraction. Among the monoterpenes, *trans*- β -ocimene, β -ocimene, and β -linalool were highly inducible during flower maturation (#9, 10, and 13, Figure 1B). Additionally, GC-MS analysis identified several sesquiterpenes or sesquiterpene alcohols that decreased during flower maturation such as such as δ -elemene (#46), α -ylangene (#47), γ -cadinene (#48) and α -copaene-11-ol (#49). Other aromatic compounds such as benzenoid/phenolpropanoid and volatile fatty acids were almost exclusively found in mature yellow flowers (Figure 1B, Figure 9, and Table 2). GC- MS analysis of flowers at night did not show any change in VOC profile suggesting that there is no diurnal changes in the emission pattern (Figure 12).

EXAMPLE 3

RNA Sequencing, *de novo* Assembly and Annotation of Transcriptome

[0102] To profile ylang ylang floral transcriptome, RNA-seq libraries that were synthesized from the ylang ylang mature yellow flowers were sequenced. Illumina sequencing runs generated more than 110 million reads of 101 base pairs (bp) and the quality of reads were evaluated by FastQC (Figure 13). Due to the absence of reference genomic sequences of ylang ylang, the Trinity method was used for *de novo* assembly of short sequence reads (Grabherr et

al., 2011). These assemblies generated a total of 45,379 non-redundant (nr) unigenes with a N50 value of 2,016 bp (Table 3). The assembled unigenes were blasted against the National Centre for Biotechnology Information (NCBI) nr protein database and protein databases from *Arabidopsis thaliana*, *Vitis vinifera*, and *Oryza sativa*. Among 45,379 non-redundant unigenes, 30,539 (67.3%) unigenes were annotated through BLASTX search with E-value $\leq 1e-3$ (Table 3). Functional classifications of Gene Ontology (GO) term of all unigenes were performed using Trinotate (Quevillon et al., 2005). Figure 2 shows enriched GO terms for the top 1,000 highly expressed transcripts. From the annotated unigenes, 16 of them were identified as TPSs, which were more than 500 bp in length.

TABLE 3
Overview of the Assembly Results of RNA-seq

# isoforms	N50 (bp)	# unigenes	# annotation	% annotation
86,512	2,016	45,379	30,539	67.3

EXAMPLE 4

RNA-seq Analysis of Different Biosynthetic Pathways Active in Flowers

[0103] GC-MS analysis of the mature flowers showed a domination of VOCs by terpenes and minor benzenoids/phenolpropanoid compounds. The RNA-seq data was used to analyze the expression profile of the precursor pathways leading to the formation of VOCs in mature flowers. Majority of ylang ylang orthologue unigenes were full length and showed high sequence similarity with known enzymes of these pathways from other plants (Figures 14A-14D). The MEP and MVA pathways are the pathways leading to the formation of mono- and sesquiterpenes. Transcripts of all the enzyme genes involved in these 2 pathways were detected in our RNA-seq data and the expression was validated by qRT-PCR (Figures 3A-3D). Additionally, the expressions of these enzyme genes were examined at earlier stages of flower development and leaves by qRT-PCR. Genes encoding the MEP and MVA enzymes were active in all stages of flower development consistent with the high production of terpenes in the flowers (Figures 3A and 3B). It has been reported that 1-deoxy-D-xylulose-5-phosphate synthase (DXS), the first enzyme of the MEP pathway is important for the overall regulation of the pathway and is encoded by a small gene family (Cordoba et al., 2011). From the RNA-seq data four different 1-deoxy-D-xylulose-5-phosphate synthase (DXS) unigenes showing different

levels of abundance in flowers and leaf were identified. One of them DXS3 belongs to clade 2, which may be related to secondary metabolism (Walter et al., 2002; Phillips et al., 2007) was highly induced in stage III flowers (Figure 3A, Figure 15, and Figure 16). The majority of the genes except *PAL* for the shikimate pathway enzymes, which produce the aromatic precursor phenylalanine for the production of benzonoids and phenylproponoids, appeared to be more expressed in mature flowers (Figures 3C and 3D). The significant benzenoids in VOCs were benzyl benzoate and benzyl salicylate and the phenylproponoid was eugenol. These compounds were only detected in mature flowers (#18, 39, and 45, Figure 1B and Table 2). Previous *in vitro* experiments have indicated that benzyl benzoate might play a role in pollinator attraction (Hoballah et al., 2005; Huber et al., 2005) or in plant defence (miticide) (Harju et al., 2004).

EXAMPLE 5

Phylogenetic Analysis of *TPS* Genes from Dwarf Ylang Ylang Flowers

[0104] Domination of terpenes is observed in the floral VOCs. From the dwarf ylang ylang RNA-seq data 4 full-length open reading frames (ORFs) of *TPS*s were PCR-amplified from cDNA pools of dwarf ylang ylang flowers. Phylogenetic analysis based on the deduced amino acid sequences of four *CoTPS* cDNAs showed that CoTPS2 (561 amino acids) belongs to the TPS-a subfamily representing the sesqui-TPSs, whereas CoTPS1 (590 amino acids) and CoTPS3 (547 amino acids) fall into the TPS-b subfamily, which consist mainly of mono-TPSs (Chen et al., 2011; Figure 4A). CoTPS4 (586 amino acids) is a member of the TPS-g, subfamily that lacks the R(R)₈W (SEQ ID NO:1) motif in the N-terminal region of mono-TPSs and produces an acyclic monoterpenes (Dudareva et al., 2003; Chen et al., 2011; Yuan et al., 2008; Figure 4A). All four CoTPSs had the conserved aspartate-rich motif (DDXXD; SEQ ID NO:2) and NSE/DTE (SEQ ID NO:3/SEQ ID NO:4) motifs that chelate the divalent metal ion, typically Mg²⁺, in C-terminal domain (Figure 4B). Both motifs are required for cyclization of the universal acyclic terpene precursors, such as genanyl and farnesyl diphosphate to synthesize mono- and sesqui-terpene, respectively (Chen et al., 2011). The arginine-tryptophan motif, R(R)₈W (SEQ ID NO:1) present at the N-terminal of most mono-TPS and in some sesqui-TPS and di-TPS were found in CoTPS1, CoTPS2, and CoTPS3 except CoTPS4 (Figure 4B). One of the distinguishing structural features between mono- and sesqui-TPS is the presence of a N-terminal plastid transit peptide (Tp) sequence. Using the signal sequence analysis programs, ChloroP ([http colon slash slash cbs dot dtu dot dk slash services slash ChloroP](http://colonslashslashcbsdotdtu dot dk slash services slash ChloroP)) and WoLF PSORT ([http colon slash slash wolffpsort dot org](http://colonslashslashwolffpsort dot org)), a putative N-terminal plastid Tp sequence of

41 and 35 amino acids for CoTPS1 and CoTPS4, respectively, was predicted indicating they are likely to be mono-TPSs. However, a putative plastid Tp sequence for CoTPS3, which was supposed to be mono-TPS belonging to TPS-b subfamily, could not be found. CoTPS2 did not contain a plastid Tp sequence which correlated well with the prediction of it being a sesqui-TPS.

EXAMPLE 6

Subcellular Localization and Expression of Four *CoTPSs*

[0105] Apart from the phylogenetic analysis and bioinformatics-based attempts to classify TPSs, their subcellular localization is also important for function prediction. This is especially true for CoTPS3, since its function was unpredictable from the bioinformatics analyses based on amino acid sequences. To address this issue, full-length cDNA of each CoTPS fused to yellow fluorescent protein (YFP) reporter gene to produce a CoTPS-YFP fusion protein was transiently expressed in *N. benthamiana* leaves using *Agrobacterium*-mediated infiltration. Figure 5 shows that CoTPS1-YFP and CoTPS4-YFP which had N-terminal plastid Tp sequence were localized in chloroplasts as expected, whereas CoTPS2-YFP and CoTPS3-YFP were distributed throughout the cytosol. Based on results of subcellular localization experiments, it is likely that CoTPS1 and CoTPS4 are involved in monoterpenes synthesis in plastids, whereas, CoTPS2 and CoTPS3 produce sesquiterpenes in the cytosol.

[0106] Transcript levels for the four *CoTPS* genes at different developmental stages of dwarf ylang ylang flowers were examined by qRT-PCR. The expression levels of all four transcripts were very low or below detection limits in leaf tissues and greatly elevated in flower tissues (Figure 6). Transcripts for three *TPS* genes, *CoTPS1*, *CoTPS3* and *CoTPS4* were highest in mature green flowers (II), whereas *CoTPS2* was highly expressed at the floral bud stage (Figure 6).

EXAMPLE 7

Functional Characterization of CoTPSs

[0107] The subcellular localization of each CoTPS-YFP provided preliminary evidence to elucidate the function of each TPS. The exact functional annotation of a new TPS requires activity characterization of the recombinant protein. To determine the enzymatic activity of CoTPSs *in vitro*, 6His-tagged CoTPSs recombinant proteins purified from *E. coli* BL21(DE3) (Figure 18) were used for *in vitro* assays. GPP (C10) or FPP (C15) was used as the common substrate for mono- and sesqui-TPS, respectively. Control assays using heat-inactivated

recombinant 6His-tagged CoTPSs did not form any terpenes from GPP or FPP (Figure 19). Figure 7A shows that CoTPS1, a member of the TPS-b family synthesized two products corresponding to β -thujene, sabinene, β -pinene and α -terpinene from GPP, but not from FPP, which were found in the essential oil profiles of dwarf ylang ylang flowers (#1, 4, 5 and 7, Figure 1 and Table 2). These results suggest that CoTPS1 is a multifunctional β -thujene/sabinene/ β -pinene/ α -terpinene synthase that is able to catalyse the synthesis of a mixture of monoterpenes namely β -thujene, sabinene, β -pinene and α -terpinene. It is not surprising because several multiproduct mono-TPSs that produce similar compounds, such as α -thujene, sabinene, α/β -pinene, α/γ -terpinene have been widely reported in other plant species (Chen et al., 2003; Fährnrich et al., 2011; Fäldt et al., 2003; Shimada et al., 2004; Lückner et al., 2002).

[0108] Similarly, recombinant CoTPS2 catalyzed the synthesis of three compounds, β -ylangene, β -copaene, and β -cubebene from FPP. These results suggest that CoTPS2 is a multifunctional β -ylangene/ β -copaene/ β -cubebene synthase capable of producing three sesquiterpenes, β -ylangene, β -copaene, and β -cubebene (Figure 7B). This finding is not unique to CoTPS2 since many TPSs are known to be multi-functional (Steele et al., 1998; Lee and Chappell, 2008). However, TPSs that produce β -ylangene/ β -copaene/ β -cubebene have not yet been reported. Of these three sesquiterpene compounds, we could detect β -ylangene and β -cubebene in the flowers of dwarf ylang ylang (#23 and 27, Figures 1 and 1B and Table 2).

[0109] CoTPS3 is a member of TPS-b family with an unusual feature that it lacks a putative N-terminal Tp sequence. Enzyme assays showed that CoTPS3 catalysed the formation of α -bergamotene from FPP (Figure 7C), which is a major sesquiterpene produced in the flowers of ylang ylang (#30, Figure 1B and Table 2). As a member of the TPS-g family, CoTPS4 was capable of utilizing GPP to synthesize an acyclic monoterpene, geraniol (Figure 7D). It was confirmed by comparison of retention time and mass spectra to those of authentic standard (Figure 19). This result was expected since the protein showed the highest amino acid identity with geraniol synthases (84%) from Madagascar periwinkle (*Catharanthus roseus*) (Figure 17D; Simkin et al., 2013). But geraniol was not detected in our GC-MS analysis of ylang ylang flowers.

EXAMPLE 8

Functional Characterization of CoTPS *in vivo*

[0110] Whether CoTPSs would produce the same terpene products *in vivo* was investigated using *Agrobacterium*-mediated transient gene expression in tobacco leaves. The YFP-fused

CoTPS1, CoTPS2, CoTPS3, or CoTPS4 was expressed in *N. benthamiana* with or without co-expression of the *Arabidopsis* 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase (HMGR). HMGR catalyses a rate limiting step of the mevalonate pathway and its overexpression is known to increase heterologous sesquiterpene production (Song et al., 2012; Jin et al., 2014). Analysis by GC-MS showed that the *in vivo* results were consistent with those obtained *in vitro*. CoTPS2-YFP produced clearly three compounds, β -ylangene, β -copaene and β -cubebene when it was co-expressed with AtHMGR (Figure 8A), whereas Co-TPS3 produced α -bergamotene when co-expressed with AtHMGR (Figure 8B). The expression of CoTPS2 or CoTPS3 alone without AtHMGR in *N. benthamiana* did not produce any terpenes, which may be due to limiting amounts of the substrate, FPP (Figure 20). CoTPS1 and CoTPS4 characterized as mono-TPS *in vitro* failed to produce any new peaks *in planta* suggesting that they might require the co-expression of additional genes, probably a rate-limiting step of the non-mevalonate pathway. Alternatively, compounds formed by these TPSs might be further metabolised endogenously by the plants.

EXAMPLE 9

Dwarf Ylang Ylang Essential Oils are Quantitatively Dominated by Sesquiterpene Compounds

[0111] Like jasmine and rose extracts, scent extracts derived ylang ylang flowers are widely used in the perfumery, cosmetic, and food industries. Plants use the vibrant colors and VOCs of flowers to attract pollinators. The chemical composition of floral VOCs produced by ylang ylang varieties has been previously reported (Gaydou et al, 1986; Benini et al., 2010, 2012; Brokl et al., 2013). These papers show the presence of volatile terpenes, benzenoid and phenylpropanoids in floral VOCs. A previous report described the composition of essential oils of ylang ylang flowers originating from Madagascar (*Cananga odorata* Hook Fil. et Thomson forma genuina) (Gaydou et al, 1986). These authors found that the primary component was the monoterpene linalool (19%), and the other major compounds were two sesquiterpenes, β -caryophyllene (10.7%) and germacrene D (10.3%) (Gaydou et al., 1986). Additionally, the variety of ylang ylang from Madagascar contained more than 20% of other aromatic compounds such as p-methylanisole, benzyl benzoate, methyl benzoate and benzyl salicylate (Gaydou et al., 1986). By contrast, our analysis on volatile essential oils of dwarf ylang ylang flowers (*C. odorata* var. fruticosa) showed that over 90% of it is composed of sesquiterpenes, such as α -farnesene (31.50%), α -bergamotene (26.79%), germacrene D (13.26%), β -caryophyllene

(11.57%), humulene (1.63%), farnesol (0.75%), *trans*- β -farnesene (0.77%) and β -ylangen (0.63%) (Figures 1A and 1B and Table 2). In addition, we also detected other groups of aromatic compounds such as benzyl benzoate, benzyl salicylate and eugenol from the flowers of dwarf ylang ylang, but they constituted less than 3% (Table 2). In the variety studied here, a dominance of α -farnesene (31.50%) which was absent from Madagascar ylang ylang and α -bergamotene (26.79%) one of the main constituents of sandalwood oil (Jones et al., 2011) was found. The differences in the chemical composition of essential oils may be due to differences in genetic background, geographic location, growth conditions and extraction methods (Benini et al., 2012; Brokl et al., 2013). Fragrant flowers from champak (*Michelia champaca* L.), indian cork (*Millingtonia hortensis* L.), and jasmine (*Jasminum sambac* L.) produce high amounts of four sesquiterpenes, β -caryophyllene, β -bergamotene, α -cubebene and β -cubebene (Samakradhamrongthai et al, 2009). Therefore, these sesquiterpenes are important contributors to typical fragrances of these flowers. In addition to α -farnesene we detected high levels of β -caryophyllene and α -bergamotene in our ylang ylang variety as well.

EXAMPLE 10

RNA-seq Uncovers Terpene Synthase Genes from Dwarf Ylang Ylang

[0112] The RNA-seq approach provided a rich resource to identify and functionally characterize *TPSs* from the flowers of dwarf ylang ylang. Approximately 16 candidate *TPS* transcripts for various mono- and sesquiterpenes were found from the transcriptome data of dwarf ylang ylang flowers. However, many of the candidate *TPS* transcripts contained partial mRNA sequences from the RNA-seq data. Amongst the four *CoTPSs* transcripts studied, the expression level of *CoTPS1* was the highest, followed by *CoTPS2* and *CoTPS3*, with *CoTPS4* being the lowest. However, the RNA-seq expression level of *CoTPSs* transcripts did not exactly correlate with the abundance of terpenes produced by these *TPSs* when analyzed by GC-MS. This may be due to post-translational modification or a reflection of different enzyme activities.

EXAMPLE 11

Sequence Characteristics of CoTPSs

[0113] According to the phylogenetic analysis, *CoTPS1* and *CoTPS3* were grouped into *TPS-b* subfamily, which commonly represents angiosperm mono-*TPSs*. Generally, the *TPS-b* group contains two distinct structural domains, the plastid TP domain and the R(R)X₈W (SEQ ID NO:1) motif for monoterpene cyclization located at N-terminal region of mature *TPS*

(Bohlmann et al., 1998). CoTPS1 appears to be a typical member of TPS-b subfamily, and has both these mono-TPS characteristics. *In vitro* studies also showed that CoTPS1 catalyzed the formation of the monoterpene, β -thujene/sabinene/ β -pinene/ α -terpinene from GPP. In contrast to CoTPS1, CoTPS3 has a conserved R(R)X₈W (SEQ ID NO:1) motif, but no TP sequence for plastid targeting, which explains its cytosolic localization in transient expression assays (Figure 5). Moreover, CoTPS3 used FPP to synthesize the sesquiterpene, α -bergamotene *in vitro* and produced same product *in vivo* (Figure 7C and Figure 8B). Hence it is a unique sesqui-TPS that contains not only monoterpene characteristic, R(R)X₈W (SEQ ID NO:1), but it also belongs to the TPS-b subfamily associated with mono-TPSs. The protein encoded by *CoTPS3* has low level of sequence identity of 45% to the α -terpineol synthase from Magnolia flower (*Magnolia grandiflora*), which is a mono-TPS most similar to CoTPS3 (Lee and Charpell, 2008). Similar sesqui-TPSs that reside in the TPS-b phylogenetic clade have been reported in tomato (*Solanum lycopersicum*) and sandalwood (*Santalum album*) (Falara et al., 2011; Jones et al., 2011).

[0114] Conversely, the CoTPS4 contains N-terminal 35 amino acid residues of a putative plastid TP sequence, but lacks the R(R)X₈W (SEQ ID NO:1) motif, a characteristic feature of TPS-b mono-TPS. CoTPS4 was annotated as a plastid geraniol synthase through Blast X analysis since it closely resembled the geraniol synthase from Madagascar periwinkle (*Catharanthus roseus*), with an amino acid sequence identity of 84% (Simkin et al., 2013). As expected, this protein catalysed the synthesis of geraniol from GPP *in vitro*. Since the new TPS-g family lacking the R(R)X₈W (SEQ ID NO:1) motif was first defined from Snapdragon TPS genes related to floral scent biosynthesis (Dudareva et al., 2003), two additional TPS genes belonging to the TPS-g family were subsequently identified from Arabidopsis and rice. These TPSs produce an acyclic monoterpene linalool (Chen et al., 2011; Yuan et al., 2008), which eventually became a prominent feature for members of the TPS-g group. As expected from a member of TPS-g family CoTPS4 produced an acyclic geraniol and the protein sequence clustered closely with the grapevine geraniol synthases of the TPS-g subfamily (Martin et al., 2010) indicating these TPS functions are highly conserved among plants.

EXAMPLE 12

CoTPS2 is a Multifunctional and Novel Sesquiterpene Synthase

[0115] Many TPSs are known to synthesize several products simultaneously. Typical multiproduct mono-TPSs such as cineole synthases, terpinene synthases, terpinolene synthases, bornyl diphosphate synthases, carene synthases and myrcene synthases additionally produce the

same compounds such as sabinene, α/β -pinene (Fähnrich et al., 2011; Chen et al., 2003; Fäldt et al., 2003; Shimada et al., 2004; Lückner et al., 2002). Most of the multiproduct TPSs are likely to synthesize one or two compounds dominantly as major products and others as minor components. Interestingly, CoTPS1 was capable of producing all four monoterpenes, β -thujene, sabinene, β -pinene, and α -terpinene at similar levels, which was named as β -thujene/sabinene/ β -pinene/ α -terpinene synthase in this study. Monoterpene thujene, usually referred to as α -thujene, has two double-bond isomers known as β -thujene and sabinene. Since all thujene synthases identified catalyses the α form, CoTPS1 possessed the ability to cyclize GPP to β -thujene, found in dwarf ylang ylang flowers.

[0116] Some sesqui-TPSs belonging to the TPS-a subfamily from *Magnolia* and kiwifruit preserved the N-terminal R(R) X_8 W (SEQ ID NO:1) motif (Lee and Chappell, 2008; Nieuwenhuizen et al., 2009). This is the case with CoTPS2 in this study. While CoTPS3 and CoTPS4 produced a single terpene product, α -bergamotene and geraniol, respectively, CoTPS2 was able to synthesize three kinds of sesquiterpenes, β -ylangene, β -copaene, and β -cubebene (Figures 7A-7D and Figures 8A and 8B). Similar to CoTPS2, many sesqui-TPSs derived from different plant species have been documented to produce multiple products which normally arise from enantiomers or common intermediates (Munck and Croteau, 1990, Steele et al, 1998, Mercke et al, 1999; Lee and Chappell, 2008).

[0117] Almost all the terpenes produced by the four CoTPSs are accounted for in the ylang ylang essential oils composition as shown in Figure 1 and Table 2, as well from other sources (Gaydou et al, 1986; Brokl et al, 2013). The two undetectable compounds, geraniol and β -cubebene may be produced in extremely small quantities and possibly can be detected by improved analytical technology like two-dimensional GC coupled to time-of-flight MS. However, it remains to be clarified if these two compounds are indeed constituents of dwarf ylang ylang floral VOCs.

[0118] TPSs for α -bergamotene and geraniol have been reported in other plant species (Lu et al., 2002; Landmann et al, 2007; Iijima et al, 2004). However, exclusive β -ylangene/ β -copaene/ β -cubebene synthase has not yet been reported. β -cubebene synthase gene has been identified in *Magnolia grandiflora* (Lee and Chappell, 2008), as Mg25, sharing 55% amino acid sequence identity and 72% similarity to CoTPS2. β -ylangene/ β -cubebene or β -copaene/ β -cubebene were found as minor peaks out of total 52 or 15 sesquiterpenes synthesized *in vitro* assays in *Abies grandis* or *Medicago truncatula* (Steele et al., 1998; Arimura et al., 2008). Interestingly, a fungal (*Coprinus cinereus*) sesqui-TPS that synthesizes 10 different

sesquiterpenes with δ -cadinene and β -copaene as the major products was capable of synthesizing β -ylangene, when the amino acid residues which presumably interact with a conserved Asp in the two metal-binding motifs were mutated (Lopez-Gallego et al, 2010). β -ylangene is a diastereomer of β -copaene, however it was not produced by the wild type sesqui-TPS (Lopez-Gallego et al., 2010). In conclusion, CoTPS2 is a multifunctional and novel TPS producing three sesquiterpenes, β -ylangene, β -copaene, and β -cubebene *in vitro* as well as *in vivo*. α -copaene, the isomer of β -copaene is a potent attractant for an agricultural pest, Mediterranean fruit flies, *Ceratitis capitata* (Nishida et al., 2000).

[0119] The use of the terms “a” and “an” and “the” and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms “comprising,” “having,” “including,” and “containing” are to be construed as open-ended terms (i.e., meaning “including, but not limited to,”) unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0120] Embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

BIBLIOGRAPHY

- [0121] **Arimura G, Garms S, Maffei M, Bossi S, Schulze B, Leitner M, Mithöfer A, Boland W.** 2008. Herbivore-induced terpenoid emission in *Medicago truncatula*: concerted action of jasmonate, ethylene and calcium signaling. *Planta* **227**, 453-464.
- [0122] **Benini C, Danflous JP, Wathelet JP, du Jardin P, Fauconnier ML.** 2010. Ylang-ylang [*Cananga odorata* (Lam.) Hook. f. & Thomson]: An unknown essential oil plant in an endangered sector. *Biotechnology, Agronomy, Society and Environment* **14**, 693-705.
- [0123] **Benini C, Mahy G, Bizoux JP, Wathelet JP, du Jardin P, Brostaux Y, Fauconnier ML.** 2012. Comparative chemical and molecular variability of *Cananga odorata* (Lam.) Hook.f. & Thomson forma genuina (ylang-ylang) in the Western Indian Ocean Islands: implication for valorization. *Chemistry & Biodiversity* **9**, 1389-1402.
- [0124] **Bohlmann J, Meyer-Gauen G, Croteau R.** 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences, USA* **95**, 4126-4133.
- [0125] **Bonfim, K. et al.** 2007. RNAi-mediated resistance to Bean golden mosaic virus in genetically engineered common bean (*Phaseolus vulgaris*). *Mol Plant Microbe Interact* **20**:717-726.
- [0126] **Brokl M, Fauconnier ML, Benini C, Lognay G, du Jardin P, Focant JF.** 2013. Improvement of Ylang-Ylang Essential Oil Characterization by GC×GC-TOFMS. *Molécules* **18**, 1783-1797.
- [0127] **Burdock GA, Carabin LG.** 2008. Safety assessment of ylang-ylang oil as a food ingredient. *Food and Chemical Toxicology* **46**, 433-445.
- [0128] **Chang YT, Chu FH.** 2011. Molecular cloning and characterization of monoterpene synthases from *Litsea cubeba*. *Tree Genetics and Genomes* **7**, 835-844.
- [0129] **Chen F, Tholl D, Bohlmann J, Pichersky E.** 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* **66**, 212-229.
- [0130] **Chiu, W. et al.** 1996. Engineered GFP as a vital reporter in plants. *Current Biology* **6**:325-330.
- [0131] **Christensen, A. H. and Quail, P. H.** 1989. Sequence analysis and transcriptional regulation by heat shock of polyubiquitin transcripts from maize. *Plant Mol Biol* **12**:619-632.
- [0132] **Christensen, A. H. et al.** 1992. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol Biol* **18**:675-689.
- [0133] **Cordoba E, Porta H, Arroyo A, San Roman C, Medina L, Rodriguez-Concepcion M, Leon P.** 2011. Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *Journal of Experimental Botany* **62**, 2023-2038.
- [0134] **Degenhardt J, Köllner TG, Gershenzon J.** 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **70**, 1621-1637.
- [0135] **De Wet, J.R. et al.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. *Mol Cell Biol* **7**:725-737.

- [0136] **Dudareva N, Cseke L, Blanc VM, Pichersky E.** 1996. Evolution of floral scent in *Clarkia*: novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *The Plant Cell* **8**, 1137-1148.
- [0137] **Dudareva N, Murfitt LM, Mann CJ, Gorenstein N, Kolosova N, Kish CM, Bonham C, Wood K.** 2000. Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. *The Plant Cell* **12**, 949-961. \
- [0138] **Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Faldt J, Miller B, Bohlmann J.** 2003. (E)- β -Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *The Plant Cell* **15**, 1227-1241.
- [0139] **Dudareva N, Klempien A, Muhlemann JK, Kaplan I.** 2013. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist* **198**, 16-32.
- [0140] **Falara V, Akhtar TA, Nguyen TT, et al.** 2011. The tomato terpene synthase gene family. *Plant Physiology* **157**, 770-789.
- [0141] **Fuentes, A. et al.** 2006. Intron-hairpin RNA derived from replication associated protein C1 gene confers immunity to tomato yellow leaf curl virus infection in transgenic tomato plants. *Transgenic Res* **15**:291-304.
- [0142] **Gaydou EM, Randriamiharisoa R, Bianchini JP.** 1986. Composition of the essential oil of Ylang-Ylang (*Cananga odorata* Hook Fil. et Thomson forma genuina) from Madagascar. *Journal of Agricultural and Food Chemistry* **34**, 481-487.
- [0143] **Goff, S.A. et al.** 1990. Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues. *EMBO J* **9**:2517-2522.
- [0144] **Grabherr MG, Haas BJ, Yassour M, et al.** 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652.
- [0145] **Guo, H.S. et al.** 2003. A chemical-regulated inducible RNAi system in plants. *Plant J* **34**:383-392.
- [0146] **Harju AT, Pennanen SM, Liesivuori J.** 2004. The efficacy of benzyl benzoate sprays in killing the storage mite *Tyrophagus putrescentiae* (Acari: Acaridae). *Annals of Agricultural and Environmental Medicine* **11**, 115-119.
- [0147] **Hoballah ME, Stuurman J, Turlings TC, Guerin PM, Connétable S, Kuhlemeier C.** 2005. The composition and timing of flower odour emission by wild *Petunia axillaris* coincide with the antennal perception and nocturnal activity of the pollinator *Manduca sexta*. *Planta* **222**, 141-150.
- [0148] **Huber FK, Kaiser R, Sauter W, Schiestl FP.** 2005. Floral scent emission and pollinator attraction in two species of *Gymnadenia* (Orchidaceae). *Oecologia* **142**, 564-575.
- [0149] **Iijima Y, Gang DR, Fridman E, Lewinsohn E, Pichersky E.** 2004. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiology* **134**, 370-379.
- [0150] **Jang IC, Yang JY, Seo HS, Chua NH.** 2005. HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signalling. *Genes and Development* **19**, 593-602.
- [0151] **Jefferson RA, Kavanagh TA, Bevan MW.** 1987. GUS fusion: β -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6**, 3901-39.

- [0152] Jin J, Panicker D, Wang Q, Kim MJ, Liu J, Yin J-L, Wong L, Jang IC, Chua NH, Sarojam R. 2014. Next generation sequencing unravels the biosynthetic ability of Spearmint (*Mentha spicata*) peltate glandular trichomes through comparative transcriptomics. *BMC genomics* **14**, 292.
- [0153] Jones CG, Moniodis J, Zulak KG, Scaffidi A, Plummer JA, Ghisalberti EL, Barbour EL, Bohlmann J. 2011. Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b subfamilies, including santalene synthases. *Journal of Biological Chemistry* **286**, 17445-17454.
- [0154] Kain, S.R. et al. 1995. Green fluorescent protein as a reporter of gene expression and protein localization. *Biotechniques* **19**:650-655.
- [0155] Kessler A, Baldwin IT. 2001. Defensive function of herbivore-induced plant volatile emissions in nature. *Science* **291**, 2104-2105.
- [0156] Knudsen JT, Tollsten L, Bergström LG. 1993. Floral scents-a checklist of volatile compounds isolated by head-space techniques. *Phytochemistry* **33**, 253-280.
- [0157] Knudsen JT, Eriksson R, Gershenzon J, Stahl B. 2006. Diversity and distribution of floral scent. *The Botanical Review* **72**, 1-120.
- [0158] Kramer W, Fritz HJ. 1987. Oligonucleotide-directed construction of mutations via gapped duplex DNA. *Methods Enzymology* **154**, 350-367.
- [0159] Landmann C, Fink B, Festner M, Dregus M, Engel KH, Schwab W. 2007. Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). *Archives of Biochemistry and Biophysics* **465**, 417-429.
- [0160] Last, D. I. et al. 1991. pEmu: an improved promoter for gene expression in cereal cells. *Theor Appl Genet* **81**:581-588.
- [0161] Lee S, Chappell J. 2008. Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. *Plant Physiology*. **147**, 1017-1033.
- [0162] Lopez-Gallego F, Waerzyn GT, Schmidt-Dannert C. 2010. Selectivity of fungal sesquiterpene synthases: Role of the active site's H-1 α Loop in Catalysis. *Applied Environmental Microbiology* **76**, 7723-7733.
- [0163] Martin DM, Fäldt J, Bohlmann J. 2004. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiology* **135**, 1908-1927.
- [0164] Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann, J. 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biology* **10**, 226.
- [0165] McElroy, D. et al. 1990. Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* **2**:163-171.
- [0166] McGarvey DJ, Croteau R. 1995. Terpenoid metabolism. *The Plant Cell* **7**, 1015-1026.
- [0167] Mercke P, Crock J, Croteau R, Brodelius PE. 1999. Cloning, expression, and characterization of epi-cedrol synthase, a sesquiterpene cyclase from *Artemisia annua* L. *Archives of Biochemistry and Biophysics* **369**, 213-222.

- [0168] Muhlemann JK, Klempien A, Dudareva N. 2014. Floral volatiles: from biosynthesis to function. *Plant, Cell & Environment* **37**, 1936-1949.
- [0169] Munck SL, Croteau R. 1990. Purification and characterization of the sesquiterpene cyclase patchoulol synthase from *Pogostemon cablin*. *Archives of Biochemistry and Biophysics* **282**, 58-64.
- [0170] Mysara, M. et al. 2011. MysiRNA-designer: a workflow for efficient siRNA design. *PLOS one* **6**(10):e25642.
- [0171] Nieuwenhuizen NJ, Wang MY, Matich AJ, Green SA, Chen X, Yauk YK, Beuning LL, Nagegowda DA, Dudareva N, Atkinson RG. 2009. Two terpene synthases are responsible for the major sesquiterpenes emitted from the flowers of kiwifruit (*Actinidia deliciosa*). *Journal of Experimental Botany* **60**, 3203-3219.
- [0172] Nishida R., Shelly TE, Whittier TS, Kaneshiro KY. 2000. α -Copaene, A Potential Rendezvous Cue for the Mediterranean Fruit Fly, *Ceratitis Capitata*? *Journal of Chemical Ecology* **26**, 87-100.
- [0173] Odell, J. T. et al. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**:810-812.
- [0174] Orlova I, Marshall-Colón A, Schnepf J, et al. 2006. Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *The Plant Cell* **18**, 3458-3475.
- [0175] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**, W116-W120.
- [0176] Phillips MA, Walter MH, Ralph SG, et al. 2007. Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpenoid resin formation of Norway spruce (*Picea abies*). *Plant Molecular Biology* **65**, 243-257.
- [0177] Samakradhamrongthai R, Utama-Ang M, Thakeow P. 2009. Identification of volatile compounds released from dry scented Thai flowers and their potential application in flower-mixed tea. *Asian Journal of Food and Agro-Industry* **2**, 525-534.
- [0178] Simkin AJ, Miettinen K, Claudel P, et al. 2013. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36-43.
- [0179] Song AA, Abdullah JO, Abdullah MP, Shafee N, Othman R, Tan EF, Noor NM, Raha AR. 2012. Overexpressing 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) in the lactococcal mevalonate pathway for heterologous plant sesquiterpene production. *PLOS ONE* **7**, e52444.
- [0180] Steele CL, Crock J, Bohlmann J, Croteau R. 1998. Sesquiterpene synthases from grand fir (*Abies grandis*): Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ -selinene synthase and γ -humulene synthase. *Journal of Biological Chemistry* **273**, 2078-2089.
- [0181] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739.

- [0182] **Vanderschuren, H. et al.** 2007a. Transgenic cassava resistance to African cassava mosaic virus is enhanced by viral DNA-A bidirectional promoter-derived siRNAs. *Plant Mol Biol* **64**:549-557.
- [0183] **Vanderschuren, H. et al.** 2007b. Engineering resistance to geminiviruses--review and perspectives. *Plant Biotechnology Journal* **5**:207-220.
- [0184] **Vogt T.** 2010. Phenylpropanoid biosynthesis. *Molecular Plant* **3**, 2-20.
- [0185] **Velten, J. et al.** 1984. Isolation of a dual plant promoter fragment from the Ti plasmid of *Agrobacterium tumefaciens*. *EMBO J* **3**:2723-2730.
- [0186] **Voinnet O, Rivas S, Mestre P, Baulcombe D.** 2003. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal* **9**, 949-956.
- [0187] **Walter M, Hans J, Strack D.** 2002. Two distantly-related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *The Plant Journal* **31**, 243-254.
- [0188] **Wang, M.B. et al.** 2000. A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus. *Mol Plant Pathol* **1**:347-356.
- [0189] **Wesley, S.V. et al.** 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J* **27**:581-590.
- [0190] **Yan, P. et al.** 2012. High-throughput construction of intron-containing hairpin RNA vectors for RNAi in plants. *PLOS one* **7**(5):e38186.
- [0191] **Yuan JS, Köllner TG, Wiggins G, Grant J, Degenhardt J, Chen F.** 2008. Molecular and genomic basis of volatile-mediated indirect defense against insects in rice. *The Plant Journal* **55**, 491-503.
- [0192] **Zhang X, Garreton V, Chua NH.** 2005. The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes and Development* **19**, 1532-1543.
- [0193] **Zrachya, A. et al.** 2007. Production of siRNA targeted against TYLCV coat protein transcripts leads to silencing of its expression and resistance to the virus. *Transgenic Res* **16**:385-398.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid encoding a protein selected from the group consisting of
 - (a) a CoTPS2 protein having the amino acid sequence set forth in SEQ ID NO:10 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:10 and having the terpene synthase activity of the CoTPS2 protein;
 - (b) a CoTPS3 protein having the amino acid sequence set forth in SEQ ID NO:6 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:6 and having the terpene synthase activity of the CoTPS3 protein;
 - (a) a CoTPS1 protein having the amino acid sequence set forth in SEQ ID NO:8 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:8 and having the terpene synthase activity of the CoTPS1 protein; and
 - (a) a CoTPS4 protein having the amino acid sequence set forth in SEQ ID NO:12 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:12 and having the terpene synthase activity of the CoTPS4 protein.
2. The isolated nucleic acid of claim 1, wherein the protein having at least 90% identity is one in which one or more of the amino acids of SEQ ID NO:10 or SEQ ID NO:6, or SEQ ID NO:8 or SEQ ID NO:12 are substituted, deleted, added, and/or inserted.
3. The isolated nucleic acid of claim 1, wherein
 - (a) the protein is CoTPS2 and the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:9;
 - (b) the protein is CoTPS3 and the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:5;
 - (c) the protein is CoTPS1 and the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:7; or
 - (d) the protein is CoTPS4 and the nucleic acid has the nucleotide sequence set forth in SEQ ID NO:11.
4. The isolated nucleic acid of any one of claims 1-3 which further comprises a plant operable promoter operably linked to the nucleic acid.

5. An expression vector comprising the isolated nucleic acid of any one of claims 1-4.
6. A transgenic plant cell, plant or plant seed comprising the isolated nucleic acid of any one of claims 1-4 stably integrated into its genome.
7. The transgenic plant cell, plant or plant seed of claim 6, wherein the plant is a member of the *Cananga* genus.
8. The transgenic plant cell, plant or plant seed of claim 6, wherein the plant is a ylang ylang (*Cananga odorata*) plant.
9. A method for producing a transgenic plant which comprises introducing the isolated nucleic acid of any one of claims 1-4 or an expression vector comprising the isolated nucleic acid of any one of claims 1-4 into a plant, wherein the transgenic plant has the nucleic acid stably integrated in its genome.
10. A method for producing a transgenic plant which comprises transfecting the isolated nucleic acid of any one of claims 1-4 or an expression vector comprising the isolated nucleic acid of any one of claims 1-4 into a plant cell or plant cells and regenerating a transgenic plant from the transfected plant cell or transfected plant cells, wherein the transgenic plant has the nucleic acid stably integrated in its genome.
11. The method of claim 9 or 10, wherein the plant is a member of the *Cananga* genus.
12. The method of claim 9 or 10, wherein the transgenic plant is a ylang ylang (*Cananga odorata*) plant.
13. A method of synthesizing at least one terpene or sesquiterpene in a plant which comprises introducing the nucleic acid of any one of claims 1-4 or an expression vector comprising the nucleic acid of any one of claims 1-4 into a plant, wherein the nucleic acid is expressed in the plant thereby synthesizing the at least one terpene or sesquiterpene.

14. A method of synthesizing at least one terpene or sesquiterpene in a plant comprising transfecting the nucleic acid of any one of claims 1-4 or an expression vector comprising the nucleic acid of any one of claims 1-4 into a plant cell or plant cells and growing a plant from the transfected plant cell or transfected plant cells, wherein the nucleic acid is expressed in the plant thereby synthesizing the at least one terpene or sesquiterpene.
15. The method of claim 13 or 14, wherein
 - (i) the at least one sesquiterpene is β -ylangene, β -copaene or β -cubebene and the nucleic acid encodes the protein of (a);
 - (ii) the at least one sequiterpene is α -bergamotene and the nucleic acid encodes the protein of (b);
 - (iii) the at least one terpene is *trans*-sabinene hydrate and the nucleic acid encodes the protein of (c); or
 - (iv) the at least one terpene is geraniol and the nucleic acid encodes the protein of (d).
16. The method of claim 13, 14 or 15, wherein the plant is a member of the *Cananga* genus.
17. The method of claim 16, wherein the transgenic plant is a ylang ylang (*Canangga odorata*) plant.
18. The method of any one of claims 13-17, wherein the nucleic acid is overexpressed.
19. An isolated protein selected from the group consisting of
 - (a) a CoTPS2 protein having the amino acid sequence set forth in SEQ ID NO:10 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:10 and having the terpene synthase activity of the CoTPS2 protein;
 - (b) a CoTPS3 protein having the amino acid sequence set forth in SEQ ID NO:6 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:6 and having the terpene synthase activity of the CoTPS3 protein;
 - (a) a CoTPS1 protein having the amino acid sequence set forth in SEQ ID NO:8 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:8 and having the terpene synthase activity of the CoTPS1 protein; and

- (a) a CoTPS4 protein having the amino acid sequence set forth in SEQ ID NO:12 or a protein having at least 90% identity to the amino acid sequence set forth in SEQ ID NO:12 and having the terpene synthase activity of the CoTPS4 protein.
20. A method of down regulating biosynthesis of at least one terpene or sesquiterpene in a plant comprising targeting expression of a *CoTPS* gene in a plant, wherein the *CoTPS* gene is selected from the group consisting of a *CoTPS2* gene, a *CoTPS3* gene, a *CoTPS1* gene and a *CoTPS4* gene.
21. The method of claim 20, wherein down regulation is achieved by introducing into a cell of the plant a nucleic acid construct that when expressed produces an RNAi molecule that suppresses or down regulates expression of a *CoTPS* gene.
22. The method of claim 20 or 21, wherein
- (i) the at least one sesquiterpene is β -ylangene, β -copaene or β -cubebene and the gene is the *CoTPS2* gene;
 - (ii) the at least one sequiterpene is α -bergamotene and the gene is the *CoTPS3* gene;
 - (iii) the at least one terpene is *trans*-sabinene hydrate and gene is the *CoTPS1* gene; or
 - (iv) the at least one terpene is geraniol and the gene is the *CoTPS4* gene.
23. A method of producing at least one terpene or sesquiterpene in a yeast cell, the method comprising exogenously expressing within the yeast cell the nucleic acid of claim 1, 2 or 3, thereby producing the at least one sesquiterpene in the yeast cell.
24. The method of claim 23, wherein
- (i) the at least one sesquiterpene is β -ylangene, β -copaene or β -cubebene and the nucleic acid encodes the protein of (a);
 - (ii) the at least one sequiterpene is α -bergamotene and the nucleic acid encodes the protein of (b);
 - (iii) the at least one terpene is *trans*-sabinene hydrate and the nucleic acid encodes the protein of (c); or

- (iv) the at least one terpene is geraniol and the nucleic acid encodes the protein of (d).