

METHODS FOR TERPENOID PRODUCTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of Singapore application No. 10201807514P, filed 31 August 2018, the contents of it being hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The invention is in the field of biotechnology. In particular, the invention relates to methods for the discovery of fungal terpene synthases and the use of fungal terpene synthases for the production of terpenoids.

BACKGROUND OF THE INVENTION

[0003] Terpenoids constitute one of the most structurally diverse classes of natural products with wide applications as pharmaceuticals (such as Taxol and artemisinin), as food coloring (such as carotenoids), flavors and fragrances (such as nootkatone and sclareol) and biofuels (such as farnesene). The terpenoid diversity is attributed primarily to terpene synthases (TPSs), which convert acyclic prenyl diphosphate precursors into a multitude of cyclic and acyclic terpene scaffolds. Specifically, the terpene skeletal diversity arises from two main features of TPSs: a large number of TPSs with vastly different functions and the ability of many TPSs to catalyze multiple terpene products from a single substrate. Almost half of the characterized monoterpene and sesquiterpene synthases produce significant amounts of additional products, apart from their main products.

[0004] Structurally, all sesquiterpene synthases (STSs) from plants, fungi and bacteria have a conserved metal binding motif (unlike plant STSs are DDXXD, fungal STSs are DE(N)XXD) and NSE triad. However, outside of the motif, there is limited sequence similarity between plant and microbial TPSs. Many plant TPSs have been discovered and studied over the past few decades, however, the study of fungal terpene synthases was lagging behind. Currently, only a handful of fungal TPS have been cloned and functionally characterized (less than 50, Table 1).

[0005] Fungi have enormous diversity (~5 million species) and outnumber plants by at least 10 times. Each fungus has an average of 10-20 putative TPS homologs in

Basidiomycota, indicating that fungal terpenoids and TPS genes represent rich but largely untapped natural resources. In addition to the discovery of novel TPS genes and terpenoids, it is also valuable to identify other potentially more efficient TPSs than existing TPSs with the same functions or products. This is due to the various applications of terpenoids and huge commercial interests in terpenoids. In some cases, the production of terpenoids was limited by insufficient activity of terpene synthases. Identification of TPS enzymes with novel products, superior activity and selectivity would greatly benefit the industrial biotechnology society for terpenoid production.

10

SUMMARY

[0006] In one aspect, there is provided a bacterial strain comprising one or more vectors encoding

- a) one or more enzymes to produce one or more terpene precursors; and
- b) a fungal terpene synthase (FTPS).

15

[0007] In another aspect, there is provided a genetically modified 1-deoxyxylulose-5-phosphate synthase (DXS) enzyme, wherein the genetic modification is a mutation at one or more amino acid positions.

[0008] In another aspect, there is provided a genetically modified fungal terpene synthase (FTPS), wherein the genetic modification is a mutation at one or more amino acid positions.

20

[0009] In another aspect, there is provided a method of producing a terpenoid comprising a) culturing the bacterial strain as described herein in an expression medium, and b) isolating the terpenoid from said expression medium.

[0010] In another aspect, there is provided a method of producing a terpenoid comprising a) culturing a bacterial strain comprising a vector encoding the genetically modified FTPS as described herein in an expression medium and b) isolating the terpenoid from said expression medium.

25

[0011] In another aspect, there is provided a fungal terpene synthase (FTPS) encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO:39.

30

[0012] In another aspect, there is provided an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO:39.

DEFINITIONS

[0013] As used herein, the term “terpene” refers to a class of organic compounds produced by plants, bacteria, fungi and insects. The building blocks of terpenes have a five-carbon isoprene unit.

[0014] As used herein, the term “terpenoid” refers to a large and diverse class of organic compounds derived from terpenes and include terpenes. The building blocks of terpenes have a five-carbon isoprene unit and contain additional functional groups, typically oxygen-containing functional groups. Terpenes are a subset of terpenoids.

10 [0015] As used herein, the term “terpene precursor” refers to a substrate that is converted to a terpene or terpenoid by a terpene synthase. The substrate may be converted to a terpene or a terpenoid.

[0016] As used herein, the term “terpene synthase” (TPS) refers to an enzyme that converts terpene precursors to terpenes and/or terpenoids. The term “fungal terpene synthase” (FTPS) refers to a terpene synthase that is isolated from a fungus.

[0017] As used herein, the term “UP” in the context of a fungal terpene synthase (FTPS) refers to a domain of the FTPS. The UP domain is situated upstream of the DW domain.

[0018] As used herein, the term “DW” in the context of a FTPS refers to a domain of the FTPS. The DW domain is situated downstream of the UP domain.

20

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

25 [0020] **Fig. 1** shows the volatile terpenoids produced by *A. aegerita*. Volatile metabolites from *A. aegerita* culture were sampled by solid phase microextraction (SPME) and analyzed by GC-MS. The structure and mass spectra of identified terpene compounds are shown in Figure 4 and 5, respectively.

[0021] **Fig. 2** shows experimental results for the improvement of the solubility and activity of DXS enzyme. In all measurements, S0 is the wide type DXS from *Escherichia coli*. (A) shows the solubility and soluble DXS at 5 h after induction. (B) shows the in vitro

30

enzymatic activity measurements of DXS and its mutants. (C) shows the specific lycopene yield when different DXS mutants were used. (D) shows the annotations for DXS mutants.

[0022] **Fig. 3** shows a schematic of an engineered *Escherichia coli* strain for screening of terpene synthases. Metabolites in the pathway are: GAP, glyceraldehyde-3-phosphate; DXP, 1-deoxy-D-xylulose-5-phosphate; MEP, methylerythritol phosphate; CDP-ME, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDPMEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate; MEC, 2-C-methyl-D-erythritol-2,4-diphosphate; HMBPP, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl pyrophosphate; FPP, farnesyl pyrophosphate and GGPP, geranylgeranyl pyrophosphate. Enzymes are: dxs, DXP synthase; dxr, DXP reductase; ispD, CDPME synthase; ispE, CDPME kinase; ispF, CDPMEP synthase; ispG, MBPP synthase; ispH, HMBPP reductase; idi, IPP isomerase; STS, sesquiterpene synthase and MTS, monoterpene synthase.

[0023] **Fig. 4** shows the terpene compounds identified in this study.

[0024] **Fig. 5** shows the mass spectra of terpene compounds analyzed in this study.

[0025] **Fig. 6** shows the predicted biosynthetic gene clusters in *A. aegerita*. Four putative TPSs (AAE3_09008, AAE3_06743, AAE3_04444 and AAE3_05024) compile to an own cluster. Four of the STS genes (AAE3_10454, AAE3_12839, AAE3_04120 and AAE3_13291) are part of clusters consists amongst other of two to five P450 monooxygenases.

[0026] **Fig. 7** shows the terpenes produced by *E. coli* expressing TPS genes from *A. aegerita*. The metabolites were analysed by GC-MS with DB5 column. The metabolite profiles analysed by GC-MS with VF-WAXms column were shown in Figure 8. Major compound peaks are labelled by numbers corresponding to structure shown below. See Figure 5 for mass spectra and Table 3 for summary of terpene compounds analysed by both DB5 and VF-WAXms column.

[0027] **Fig. 8** shows the GC-MS profiles of terpenes produced by *E. coli* expressing TPS genes from *A. aegerita*. The metabolites were analyzed by VF-WAXms column. 'Ctrl' is the GC-MS profile of volatile metabolites produced by *E. coli* strain with empty vector. Indole (*) endogenously produced by *E. coli* serves as an internal reference to compare the relative amount of the terpene production.

- [0028] **Fig. 9** shows the nuclear magnetic resonance (NMR) spectroscopy analysis of the product of AAE3_10454.
- [0029] **Fig. 10** shows the use of essential oils as chemical standards to identify terpene compounds identified. VF-WAXms column was used in this study.
- 5 [0030] **Fig. 11** shows the use of essential oils as chemical standards to identify terpene compounds identified. DB-5ms column was used in this study.
- [0031] **Fig. 12** shows the GC-MS profile of monoterpenes produced by AAE3_9164.
- [0032] **Fig. 13** shows the phylogenetics of characterized TPS homologues from four fungi. TPSs from *A. aegerita* (AAE3), *C. cinereus* (Cop), *O. olearius* (Omp), *Stereum*
- 10 *hirsutum* (Steh1) and the single TPS described from *Armillaria gallica* (Pro1). Sequences used in the final alignment can be found in Table 1. The TPSs highlighted with a square (“■”) had no detected terpene products.
- [0033] **Fig. 14** shows the comparison of different GC columns for the analysis of terpenes. Niaouli essential oil was used as standards of viridiflorene **6** and viridiflorol **7**. Viridiflorol **7**
- 15 has relatively lower signal than viridiflorene **6** in DB-5 column as confirmed by authentic chemical standards from Santa Cruz Biotechnology (Figure 17). The difference of relative intensity of viridiflorol and caryophyllene (structurally similar to viridiflorene) in DB-5 and VF-WAXms were further verified by authentic standards.
- [0034] **Fig. 15** shows the similarity analysis of AAE3_12839 (SEQ ID NO: 34) and
- 20 TPS31 (SEQ ID NO: 75) from *Solanum lycopersicum* (Tomato). The fungal viridiflorene synthase (AAE3_12839) shares little identity and similarity with plant viridiflorene synthase.
- [0035] **Fig. 16** shows the similarity analysis of AAE3_13291 (SEQ ID NO: 32) and MqTPS1 (SEQ ID NO: 76) from *Melaleuca quinquenervia*. The fungal viridiflorol synthase (AAE3_12839) shares little identity and similarity with plant viridiflorol synthase.
- 25 [0036] **Fig. 17** shows the proposed reaction mechanisms for the formation of major products. The carbocation from FPP ionization undergoes two different primary ring closures (1,10 or 1,11 closure). Major compounds produced by recombinant *A. aegerita* TPSs are labelled by numbers, including Δ^6 -protolilludene **1**, γ -muurolene **2**, β -cadinene **3**, δ -cadinene **4**, α -muurolene **5**, viridiflorene **6**, viridiflorol **7**, δ -cadinol **8** and epicubenol **10**. See Figure 5
- 30 for the mass spectra.
- [0037] **Fig. 18** shows the phylogenetic tree of TPS homologs identified in 85 Basidiomycota and 239 Ascomycota genomes. (A) shows all the fungal TPSs clustered into

seven distinct clades. The characterized TPSs in this study and in literature were labelled in the figure. See more information in Table 1. In particular, these include *A. aegerita* (AAE3), *C. cinereus* (Cop), *O. olearius* (Omp), *Stereum hirsutum* (Steh1), *Hypoxylon sp* (Hyp), *Fusarium fujikuroi* (Ffsc4, Ffsc6, STC3 and STC5) and a few aristolochene synthases (AtARS and PrARS). Most of Basidiomycota TPSs (including all the 11 *A. aegerita* TPSs) clustered in clade I, II and III, but Ascomycota TPSs scattered in clade IV, V, VI and VII. (B) shows potential Δ^6 -protoilludene synthases based on the phylogenetic analysis. The TPSs highlighted with a circle (“●”) were characterized in this study or in literature.

[0038] **Fig. 19** shows a bioinformatics-guided predictive framework – all-by-all BLAST.

10 The All-by-all BLAST of the 1408 putative TPS candidates was performed with enzyme function initiative (EFI) - enzyme similarity tool (EST). Sequence similarity networks (SSN) were generated by filtering the sequences into clusters at the alignment score of 100. SSNs were used for visualization by Cytoscape version 3.5.1. The model obtained here was used together with the phylogenetic tree in Figure 13 to predict the TPS functions based on the sequence similarity and characterized TPS in this study and in literature.

[0039] **Fig. 20** shows a predictive example of the clustering of putative fungal linalool/nerolidol synthases (LNS). Based on EFI-EST analysis, a group of TPS homologues were clustered with AAE3_9435. By setting the alignment score to between 80 and 90, a smaller set of candidates were selected.

20 [0040] **Fig. 21** shows a predictive example of the validation of putative fungal linalool/nerolidol synthases. Selected putative LNSs were expressed in the engineered *E. coli* strains. LNSs chosen here in the cluster are from *Agrocybe aegerita* (AAE3), *Agrocybe pediades* (Agrped1), *Galerina marginata* (Galma), *Hypholoma sublateritium* (Hypsu1), *Hebeloma cylindrosporum* (M413). Agrped1_689675 is found to a novel monoterpene synthase, linalool synthase (LS), while the others are bifunctional LNSs.

[0041] **Fig. 22** shows the Sequencing alignment of validated LNSs and the LS. Sequencing alignment indicated similar positions of these TPSs are 107/34= 31%. Based on the alignment, a key amino acid F204 was identified that could impact the activity of the LS (Agrped1_689675).

30 [0042] **Fig. 23** shows the results of the engineering of the LS Agrped1_689675. The three mutants (F204D, F204G and F204R) have different product profiles to that of the wildtype enzyme. It produced both geranyl acetate (predicted by National Institute of Standards and

Technology (NIST) library) and linalool. The other mutants (F204I, F204L and F204V) share the same product (linalool) with the wildtype Agrped1_689675.

[0043] **Fig. 24** shows a predictive example of the clustering and validation of putative fungal viridiflorol synthases. According to the phylogenetic tree in Figure 18 and 19, 15 fungal TPS homologs were closely clustered. Four of them (Sphst_47084 from *Sphaerobolus stellatus*, Denbi1_816208 from *Dendrothele bispora*, Galma_104215 from *Galerina marginata* and Pilcr_825684 from *Piloderma croceum*) were recombinantly expressed in *E. coli* and their products were analysed. Both phylogenetic analysis and EFI-EST analysis have very accurate prediction. The TPSs highlighted with a circle (“●”) were characterized in this study.

[0044] **Fig. 25** shows the metal binding motif of the characterized TPSs in this study and in literature.

[0045] **Fig. 26** shows crystal structures of homologue models for Agrped1_689675 and Agrped1_689675, where the substrate-binding pockets are highlighted. The model was generated using the Swiss Model homology-modeling server and alignment mode with 5nx6 and 5nx5 as templates. Protein models were visualized and aligned with their template structure using PyMol.

[0046] **Fig. 27** shows the crystal structure of DXS (PDB ID: 2o1s), where the substrates and mutated amino acids are highlighted.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0047] In a first aspect, the present invention refers to a bacterial strain comprising one or more vectors encoding a) one or more enzymes to produce one or more terpene precursors, and b) a fungal terpene synthase (FTPS).

25 [0048] It will be appreciated by a person skilled in the art that the one or more vectors comprise polynucleotide sequences that encode the one or more enzymes to produce one or more terpene precursors and FTPS.

[0049] The polynucleotides encoding the one or more enzymes to produce one or more terpene precursors and the FTPS may be located on separate vectors, on a single vector or combinations thereof. In one embodiment, the polynucleotides encoding the one or more enzymes to produce one or more terpene precursors are in a single vector and the polynucleotide encoding the FTPS is in a separate vector.

[0050] In one embodiment, the one or more vectors comprise one or more nucleotide sequences encoding the one or more enzymes and the FTPS, operably linked to a promoter.

[0051] In some embodiments, the promoter is a constitutive promoter or an inducible promoter. In a preferred embodiment, the promoter is an inducible promoter. Examples of
5 inducible promoters include but are not limited to T7 RNA polymerase promoter, *araBAD* promoter, a *lac* promoter, a *trp* promoter and a *Tac* promoter (*ptac*) or the variants of these promoters.

[0052] In a preferred embodiment, the inducible promoter is T7 RNA polymerase promoter.

10 [0053] In one embodiment, the one or more enzymes to produce the one or more terpene precursors are part of the 1-deoxy-D-xylulose 5-phosphate (DXP) pathway. It will be appreciated to a person of skill in the art that the DXP pathway is also referred to as the non-mevalonate pathway, the mevalonate-independent pathway or the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP/DOXP) pathway. The DXP pathway
15 converts pyruvate and glyceraldehyde-3-phosphate to terpene precursors and the enzymes in this pathway include DOXP synthase (DXS), DXP reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (IspD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF), HMB-PP synthase (ispG), HMB-PP reductase (IspH) and isopentenyl diphosphate isomerase
20 (IDI).

[0054] In a preferred embodiment, the enzyme is 1-deoxyxylulose-5-phosphate synthase (DXS), isopentenyl diphosphate isomerase (IDI) or both.

[0055] In one embodiment, the DXS comprises the amino acid sequence set forth in SEQ ID NO: 6.

25 [0056] In some embodiments, the DXS may be genetically modified. The genetic modification may be a mutation at one or more amino acid positions of the amino acid sequence encoding the DXS. In some examples, the mutation is an amino acid substitution, insertion, deletion or combinations thereof.

[0057] In some embodiments, the genetically modified DXS has a higher solubility than
30 an unmodified or wild-type DXS.

[0058] In some embodiments, the mutation is selected from the group consisting of H105T, E210D, Q459L, L415T and a combination thereof of SEQ ID NO: 6.

- [0059] In a preferred embodiment, the mutation is H105T.
- [0060] In another preferred embodiment, the mutation is E210D, Q459L and L415T.
- [0061] In one embodiment, the genetically modified DXS comprises the amino acid sequence set forth in SEQ ID NO: 24 or 25.
- 5 [0062] In one embodiment, the DXS comprises an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequence set forth in SEQ ID NO: 6, SEQ ID NO: 24 or 25.
- [0063] In one embodiment, the one or more enzymes to produce the one or more terpene precursors is expressed at an elevated level compared to a wild type enzyme. The one or
10 more enzymes may be genetically modified.
- [0064] In some embodiments, the terpene precursors described herein is farnesyl pyrophosphate (FPP), geranyl pyrophosphate (GPP), geranylgeranyl pyrophosphate (GGPP), or combinations thereof.
- 15 [0065] In a preferred embodiment, the terpene precursors are FPP and/or GPP.
- [0066] The bacterial strain described herein comprises one more vectors encoding a fungal terpene synthase (FTPS). In a preferred embodiment, the FTPS is a monoterpene synthase or a sesquiterpene synthase. In a further preferred embodiment, the FTPS is a linalool synthase, a nerolidol synthase or a linalool and nerolidol synthase (LNS).
- 20 [0067] In some embodiments, the FTPS is isolated from *Agrocybe aegerita*, *Agrocybe pediades*, *Galerina marginata*, *Hypholoma sublateritium*, *Dendrothele bispora*, *Moniliophthora roreri*, *Piloderma croceum*, *Sphaerobolus stellatus*, *Coprinopsis cinerea*, *Omphalotus olearius*, *Fomitopsis pinicola*, *Stereum hirsutum*, *Fusarium graminearum*, *Fusarium fujikuroi*, *Fusarium sporotrichioides*, *Aspergillus terreus*, *Penicillium roqueforti*,
25 *Hypoxylon sp.*, *Armillaria gallica*, *Botrytis cinerea*, *Daldinia eschscholzii* or combinations thereof.
- [0068] In one embodiment, the FTPS is isolated from *Agrocybe aegerita*, *Agrocybe pediades*, *Galerina marginata*, *Hypholoma sublateritium*, *Hebeloma cylindrosporum* or combinations thereof.
- 30 [0069] In a preferred embodiment, the FTPS is isolated from *Agrocybe aegerita* or *Agrocybe pediades*.

[0070] In some embodiments, the FTPS comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33 and SEQ ID NO: 34.

5 [0071] In some embodiments, the FTPS comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39.

[0072] In one example, the bacterial strain described herein contains an FTPS that is expressed at a higher level than a wild-type FTPS.

10 [0073] In some embodiments, the FTPS may be genetically modified. The genetic modification may be an amino acid substitution, insertion, deletion, C-terminal truncation, N-terminal truncation or combinations thereof. The mutation may be one or more mutations in the UP domain of the FTPS, one or more mutations in the DW domain of the FTPS, or both. The UP and DW domains would be understood by the skilled person to vary based on the
15 fungal strain. In one example, the UP domain of the FTPS isolated from Agrped1_689675 (SEQ ID NO: 3) is characterized by amino acid positions 1-170. In another embodiment, the UP domain of the FTPS isolated from Agrped1_689671 (SEQ ID NO: 2) is characterized by amino acid positions 1-169. In yet another embodiment, the DW domain for the FTPS isolated from Agrped1_689675 (SEQ ID NO: 3) is characterized by amino acid positions
20 171-325. In yet another embodiment, the DW domain of the FTPS isolated from Agrped1_689671 (SEQ ID NO: 2) is characterized by amino acid positions 170-324.

[0074] In a preferred embodiment, the mutation is F204D, F204G, F204R, F204I, F204L, F204V or combinations thereof of SEQ ID NO: 3.

25 [0075] In some embodiments, the genetically modified FTPS comprises an amino acid sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO: 16 SEQ ID NO:17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22 and SEQ ID NO: 23.

30 [0076] In some embodiments, the FTPS comprises an amino acid sequence having at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to amino acid sequences set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID

NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO: 16, SEQ ID NO:17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33 and
5 SEQ ID NO: 34.

[0077] In some embodiments, the bacterial strain is modified or genetically modified.

[0078] In one embodiment, the bacterial strain described herein is *Escherichia coli*.

[0079] In one aspect, the present invention refers to a genetically modified 1-deoxyxylulose-5-phosphate synthase (DXS) enzyme, wherein the genetic modification is a
10 mutation at one or more amino acid positions. In one embodiment, the mutation described herein is an amino acid substitution or insertion or deletion. In yet another embodiment, the mutation is selected from the group consisting of H105T, E210D, Q459L, L415T and a combination thereof of SEQ ID NO: 6.

[0080] In a preferred embodiment, the mutation is E210D, Q459L and L415T.

15 [0081] In another preferred embodiment, the mutation is H105T.

[0082] In another aspect, the present invention refers to a genetically modified DXS enzyme comprising an amino acid sequence as set forth in SEQ ID NO: 24 or SEQ ID NO: 25.

[0083] In yet another aspect, the present invention refers to a genetically modified DXS
20 enzyme comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequence set forth in SEQ ID NO: 24 or SEQ ID NO: 25.

[0084] In one aspect, the present invention refers to a genetically modified fungal terpene
25 synthase (FTPS), wherein the genetic modification is a mutation at one or more amino acid positions.

[0085] The genetically modified FTPS is modified relative to a wild type or unmodified FTPS. In one embodiment, the unmodified FTPS is isolated from *Agrocybe aegerita*, *Agrocybe pediades*, *Galerina marginata*, *Hypholoma sublateritium*, *Hebeloma cylindrosporum* or combinations thereof.

30 [0086] In a preferred embodiment, the unmodified FTPS is isolated from *Agrocybe aegerita* or *Agrocybe pediades*.

[0087] In some embodiments, the mutation described herein is an amino acid substitution, insertion, deletion, C-terminal truncation, N-terminal truncation or combinations thereof. The mutation may be one or more mutations in the UP domain of the FTPS, one or more mutations in the DW domain of the FTPS, or both. The UP and DW domains would be understood by the skilled person to vary based on the fungal strain. In one example, the UP domain of the FTPS isolated from Agrped1_689675 (SEQ ID NO: 3) is characterized by amino acid positions 1-170. In another embodiment, the UP domain of the FTPS isolated from Agrped1_689671 (SEQ ID NO: 2) is characterized by amino acid positions 1-169. In yet another embodiment, the DW domain for the FTPS isolated from Agrped1_689675 (SEQ ID NO: 3) is characterized by amino acid positions 171-325. In yet another embodiment, the DW domain of the FTPS isolated from Agrped1_689671 (SEQ ID NO: 2) is characterized by amino acid positions 170-324.

[0088] In a preferred embodiment, the mutation is selected from the group consisting of F204D, F204G, F204R, F204I, F204L and F204V of SEQ ID NO: 3.

[0089] In some embodiments, the genetically modified FTPS described herein is a linalool synthase, nerolidol synthase or both.

[0090] In one embodiment, the present invention refers to a genetically modified FTPS comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequence selected from the group consisting of: SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22 and SEQ ID NO: 23.

[0091] In one aspect, the present invention refers to a method of producing a terpenoid comprising a) culturing the bacterial strain as described herein in an expression medium and b) isolating the terpenoid from said expression medium.

[0092] Culturing the bacterial strain in the expression medium will allow the expression of the one or more enzymes to produce one or more terpene precursors and expression of the FTPS.

[0093] In another aspect, the present invention refers to a method of producing a terpenoid comprising a) culturing a bacterial strain comprising a vector encoding the genetically

modified FTPS as described herein in an expression medium and b) isolating the terpenoid from said expression medium.

[0094] The expression medium may be any culture medium that supports growth of the bacterial strain. The expression medium may comprise inducers capable of inducing the inducible promoter in the one or more vectors. The expression medium may also be an auto-inducing medium. In one example, the auto-inducing medium is ZYM5052. In other examples, the auto-inducing medium is lysogeny broth (LB), Terrific Broth (TB) or 2xPY.

[0095] The expression medium may be further supplemented with spherical C18 resin or Ni-nitrilotriacetic acid resin.

[0096] In one embodiment, the method described herein further comprises the step of isolating the FTPS from the bacterial cell and mixing the isolated FTPS with one or more terpene precursors to produce the terpenoid. The isolated FTPS may be mixed with the one or more terpene precursors in the same cell culture vessel or in a different vessel from the original culture. The FTPS may be isolated using a variety of methods. In some embodiment, the FTPS is isolated the bacterial cell by Ni-nitrilotriacetic acid resin, column based methods or both.

[0097] In another embodiment, the isolated FTPS described herein is further purified prior to mixing with one or more terpene precursors.

[0098] In yet another embodiment, the isolated FTPS is further mixed with one or more additional enzymes prior to mixing with one or more terpene precursors.

[0099] In one embodiment, the one or more additional enzymes described herein is Acetyl-CoA acetyltransferase, Hydroxymethylglutaryl-CoA synthase (HMGS), Hydroxymethylglutaryl-CoA synthase reductase (HMGR), IDI, melonate kinase (MK), Phosphomevalonate kinase (PMK) or mevalonate diphosphate decarboxylase (MVD1). In some embodiments, the Acetyl-CoA acetyltransferase is PhaA or atoB. **[SF: any others?]**

[00100] In one embodiment, terpenes or terpenoids may be produced using the FTPS of the present invention as follows: 1. Using crude cell lysate, the bacterial cells expressing only FTPS are harvested and lysed by freeze/thaw method and/or sonication method. The lysed cell supernatant containing soluble FTPS is mixed with substrates (GPP or FPP), 2.5 mM MgCl₂ and 50 mM Tris/HCl buffer (pH 6.5-8.5) to produce terpenoids at 30-37°C. In another example, the FTPS will be purified from the bacterial cells by Ni-nitrilotriacetic acid resin and/or column-based method. The purified FTPS is mixed with substrates (GPP or FPP), 2.5

mM MgCl₂ and 50 mM Tris/HCl buffer (pH 6.5-8.5) to produce terpenoids at 30-37°C. In addition, the FTPS may be coupled into a multienzyme reaction, for example at pH 7.5 and at 30°C, by mixing the FTPS with other enzymes such as IDI, MK, PMK or mevalonate pyrophosphate decarboxylase (PMD) to convert mevalonate into terpenoids.

5 [00101] In one embodiment, the product of the method described herein is a monoterpenoid, sesquiterpenoid or a mixture of both. In some embodiments, the monoterpenoid is selected from the group consisting of β -myrcene, linalool, geranyl acetate and combinations thereof.

[00102] In some embodiments, the sesquiterpenoid is selected from the group consisting of
10 Δ 6-protolludene, α -muurolene, γ -muurolene, β -cadinene, β -copaene, δ -cadinene, γ -cadinene, epizonarene, α -cubebene, cubebol, epicubenol, nerolidol, viridiflorol, viridiflorene, α -cadinol, α -epi-cadinol, β -selinene, α -selinene, T-muurolol, β -elemene, β -gurjunene, germacrene A, germacrene D and combinations thereof.

[00103] In a preferred embodiment, the product is Δ 6-protolludene, linalool, geranyl
15 acetate or combinations thereof.

[00104] The present invention also discloses the use of the FTPS described herein to produce one or more terpenoids.

[00105] In some embodiments, the one or more terpenoids is selected from the group consisting of β -myrcene, linalool, geranyl acetate, Δ 6-protolludene, α -muurolene, γ -
20 muurolene, β -cadinene, β -copaene, δ -cadinene, γ -cadinene, epizonarene, α -cubebene, cubebol, epicubenol, nerolidol, viridiflorol, viridiflorene, α -cadinol, α -epi-cadinol, β -selinene, α -selinene, T-muurolol, β -elemene, β -gurjunene, germacrene A, germacrene D, trans- β -ocimene, β -cubebene, α -isocomene, longifolene, cadina-3,5-diene, caryophyllene, α -humulene, cubenene, calamenene, cubenol, δ -cadinol, cadina-1(6), 4-diene and combinations
25 thereof.

[00106] In another embodiment, the one or more terpenoids described herein is produced *in vitro* or *in vivo*. In some embodiments, the one or more terpenoids is produced *in vivo* in a bacterial cell, a yeast cell, a plant cell, an animal cell or a fungal cell. In one example, the bacterial cell is an *E.coli* cell. In another example, the yeast cell is a *Saccharomyces*
30 *cerevisiae* or a *Yarrowia lipolitica* cell.

[00107] In another embodiment, the one or more terpenoids described herein is produced *in vitro*.

[00108] In some embodiments, the FTPS described herein is isolated from a bacterial cell, a yeast cell, a plant cell, an animal cell or a fungal cell, and mixed with one or more terpene precursors to produce the one or more terpenoids.

[00109] In another embodiment, the isolated FTPS is further mixed with one or more additional enzymes prior to mixing with one or more terpene precursors. In one example, the one or more additional enzymes is IDI, MK, PMK or PMD.

[00110] The present invention also discloses a vector comprising a polynucleotide sequence encoding a 1-deoxyxylulose-5-phosphate synthase (DXS) enzyme comprising an amino acid sequence set forth in SEQ ID NO: 6, or a genetically modified 1-deoxyxylulose-5-phosphate synthase (DXS) enzyme as described herein.

[00111] In another example, the present invention refers to a vector comprising a polynucleotide sequence encoding a fungal terpene synthase (FTPS) comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:5, or a genetically modified fungal terpene synthase (FTPS) as described herein.

[00112] The polynucleotide sequences generated from amino acid sequences may be optimized for improved expression in a host cell or an expression vector. The DNA sequences may be generated from amino acid sequences to have optimised Codon Adaptation Index (>0.6) and GC percentage (40-60%). Codon usage frequency table may be based on a strain of bacterial cell, for example, on *Escherichia coli K-12 MG1655* strain. In most cases, a guided random method based on a Monte Carlo algorithm may be used. However, manual adjustments may be introduced to remove certain regions with complex secondary structures or repeated sequences. It will generally be understood that various codon optimization methods may be employed to improve expression of a protein or polypeptide in a host cell or expression vector.

[00113] In one aspect, the present invention refers to an FTPS encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO: 39.

[00114] In another aspect, the present invention refers to an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO:39.

[00115] In some embodiments, the nucleic acid sequence has at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or, at least 99% or 100% identity to the nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO: 39.

5 [00116] The bacterial strain as disclosed herein may be used to characterize a FTPS. In another embodiment, the FTPS may be characterized by a product produced by the FTPS, the activity of the FTPS, or both.

[00117] The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed
10 herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible
15 within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

20 [00118] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

25 [00119] Other embodiments are within the following claims and non-limiting examples. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

EXPERIMENTAL SECTION

[00120] Non-limiting examples of the invention and comparative examples will be further described in greater detail by reference to specific Examples, which should not be construed as in any way limiting the scope of the invention.

- 5 [00121] The fungal TPSs that have been cloned and functionally characterized are shown in Table 1.

[00122] **Table 1.** Published functionally characterized fungal terpene synthases.

	Accession or JGI Protein ID	Abbreviation	Organisms	Main product	Minor products	Reference
1	EAU89322	Cop1	Coprinopsis cinerea	germacrene A		(Agger et al., 2009)
2	EAU85264	Cop2	Coprinopsis cinerea	germacrene A		(Agger et al., 2009)
3	EAU88892	Cop3	Coprinopsis cinerea	α -muurolene	β -elemene, γ -muurolene, germacrene D and δ -cadinene	(Agger et al., 2009)
4	EAU85540	Cop4	Coprinopsis cinerea	δ -cadinene	β -cubebene, sativene, β -copaene, cubebol	(Agger et al., 2009)
5	EAU89298	Cop6	Coprinopsis cinerea	α -cuprenene		(Agger et al., 2009)
6	/	Omp1	Omphalotus olearius	α -muurolene		(Wawrzyn et al., 2012)
7	/	Omp3	Omphalotus olearius	α -muurolene	β -elemene and selina-4,7-diene	(Wawrzyn et al., 2012)
8	/	Omp4	Omphalotus olearius	δ -cadinene	16 different sesquiterpenes	(Wawrzyn et al., 2012)
9	/	Omp5	Omphalotus olearius	γ -cadinene	epi-zonarene	(Wawrzyn et al., 2012)
10	/	Omp6	Omphalotus olearius	Δ 6-protoilludene		(Wawrzyn et al., 2012)
11	/	Omp7	Omphalotus olearius	Δ 6-protoilludene		(Wawrzyn et al., 2012)
12	/	Omp9	Omphalotus olearius	α -barbatene	β -barbatene	(Wawrzyn et al., 2012)
13	/	Omp10	Omphalotus olearius	trans-dauca-4(11),8-diene	daucene	(Wawrzyn et al., 2012)
14	/	Fomp1_84944	Fomitopsis pinicola	α -cuprenene		(Wawrzyn et al., 2012)
15	/	Stehi_64702	Stereum hirsutum	Δ 6-protoilludene		(Quin et al., 2013)
16	/	Stehi_73029	Stereum hirsutum	Δ 6-protoilludene		(Quin et al., 2013)
17	/	Stehi_25180	Stereum hirsutum	Δ 6-protoilludene		(Quin et al., 2013)
18	/	Stehi_128	Stereum	δ -cadinene	β -copaene,	(Quin et al.,

8		017	hirsutum		sativene, γ -muurolene, α -muurolene etc	2013)
1 9	/	Stehi_159 379	Stereum hirsutum	β -barbatene	α -barbatene and β -barbatene	(Quin et al., 2013)
2 0	ACY69978	CLM1 FgLS	Fusarium graminearum	longiborneol		(McCormick et al., 2010)
2 1	CCP20071.1	Ffsc6	Fusarium fujikuroi	(-)- α -acorenol		(Brock et al., 2013)
2 2	CCP20072.1	Ffsc4	Fusarium fujikuroi	koraiol		(Brock et al., 2013)
2 3	AAD13657	FsTDS	Fusarium sporotrichioides	trichodiene		(Rynkiewicz et al., 2001)
2 4	AAF13264	AtARS	Aspergillus terreus	aristolochene		(Cane and Kang, 2000)
2 5	AAA33694	PrARS	Penicillium roqueforti	aristolochene		(Hohn and Plattner, 1989)
2 6	KJ433269	Hyp1	Hypoxylon sp.	trans-nerolidol		(Shaw et al., 2015)
2 7	KJ433270	Hyp2	Hypoxylon sp.	δ -cadinene	21 other peaks	(Shaw et al., 2015)
2 8	KJ433271	Hyp3	Hypoxylon sp.	1,8-cineole (C10)	D-limonene (C10)	(Shaw et al., 2015)
2 9	KJ433272	Hyp4	Hypoxylon sp.	D-limonene (C10)	12 other peaks	(Shaw et al., 2015)
3 0	KJ433273	Hyp5	Hypoxylon sp.	β -ocimene (C10)	sabinene (C10), α -bulnesene and unknown peaks	(Shaw et al., 2015)
3 1	/	Pro1	Armillaria gallica	Δ 6-protolludene		(Engels et al., 2011)
3 2	CCT65043	STC3	Fusarium fujikuroi	(+)-eremophilene		(Burkhardt et al., 2016)
3 3	CCT75704	STC5	Fusarium fujikuroi	(-)-guaia-6,10(14)- diene		(Burkhardt et al., 2016)
3 4	AAQ16575	BcBOT2 or BcPSPS	Botrytis cinerea	presilphiperfolan- 8 β -ol		(Moraga et al., 2016)
3 5	JGI ID:17536	EC12- PGS	Daldinia eschscholzii EC12	Guaiene	Pinene (C10)	(Wu et al., 2016)
3 6	JGI ID:315006	EC12-GS	Daldinia eschscholzii EC12	Gurnunene		(Wu et al., 2016)
3 7	JGI ID:24646	EC12-SS	Daldinia eschscholzii EC12	Selinene		(Wu et al., 2016)
3 8	JGI ID:70183	EC12-ILS	Daldinia eschscholzii EC12	IsoLedene		(Wu et al., 2016)
3 9	JGI ID:6706	CI4A-CS	Hypoxylon sp. CI4A	Caryophyllene		(Wu et al., 2016)
4 0	JGI ID:322581	CI4A-CPS	Hypoxylon sp. CI4A	Chamigrene	Pinene (C10)	(Wu et al., 2016)
4 1	JGI ID:397991	CO27-CS	Hypoxylon sp. CO27	Caryophyllene		(Wu et al., 2016)

4 2	JGI ID:392541	CO27- CPS	Hypoxylon sp. CO27	Chamigrene	Pinene (C10)	(Wu et al., 2016)
4 3	JGI ID:373976	EC38-CS	Hypoxylon sp. EC38	Caryophyllene		(Wu et al., 2016)
4 4	JGI ID:328361	EC38- CPS	Hypoxylon sp. EC38	Chamigrene	Pinene (C10)	(Wu et al., 2016)

[00123] *Cultivation of Agrocybe aegerita and analysis of its fruiting bodies*

[00124] *Agrocybe aegerita* wildtype-strain AAE-3 was grown at 24 °C in the dark in modified crystallizing dishes (Figure 1; lower dish: 70 mm in diameter, upper dish: 80 mm in diameter; glass pipe attached to the upper dish: outer diameter 16 mm, inner diameter 14 mm) with 16 mL 1.5% MEA (containing 15 g malt extract and 15 g agar per liter) and sealed with Parafilm. The ten days after the inoculation, the mycelium covered the complete agar surface. The Parafilm was removed and the samples were transferred to a climate chamber (24 °C, 95%rH, 12/12 h day/night rhythm) and cultured on glass plates for further 16 days. Volatile organic compounds were collected by solid phase microextraction (SPME) using a divinylbenzene-carboxen-polydimethylsiloxane (50/30 µm DVB/CAR/PDMS) fiber. Beginning with day 10 after inoculation, volatiles were absorbed directly in the crystallizing dishes for 14 h (7/7 h day/night). This extraction was carried out every second day. For GC-MS analysis an Agilent Technologies 7890A gas chromatograph (Agilent Technologies, Waldbronn, Germany) equipped with a VF WAXms column (Agilent Technologies; 30 m x 0.25 mm, 0.25 µm) and connected to an Agilent 5975C MSD Triple Axis mass spectrometer (MS) was used. Helium was used as gas carrier, with a flow rate of 1.2 ml x min⁻¹. Mass spectra were acquired in the mass range of 33 300 m/z. Ionisation was performed by electron impact at 70 eV with an ion source temperature set at 230 °C. The SPME fiber was inserted into the injector of the gas chromatograph for thermal desorption in splitless mode for 1 min, with the injector temperature held at 250 °C. The GC oven temperature was programmed to ramp from 40 °C (held for 3 min) to 240 °C (held for 7 min) at 5 °C x min⁻¹. Volatile compounds were identified by comparing mass spectra with data from the NIST14 database and matching determined retention indices with published ones. Furthermore, Cubeb oil and α humulene were used as standards.

[00125] *Gas Chromatography-Mass Spectrometry Analysis of Terpenoids*

[00126] Volatile compounds in the headspace were sampled at room temperature for 15 min by SPME with a DVB/CAR/PDMS (50/30 µm divinylbenzene/carboxen/polydimethylsiloxane) fiber (length 1 cm; Supelco, Steinheim,

Germany). Compounds were desorbed in the split/splitless inlet (250 °C or 150 °C; SPME liner, 0.75mm i.d.; Supelco) of an Agilent 7980B gas chromatography equipped with an Agilent 7200 accurate-mass quadrupole time-of-flight (GC/MS-TOF; Agilent Technologies, Singapore) for 1 min. In addition, for liquid culture analysis, dodecane (20% v/v) was used to
5 extract the terpenoid produced in *E. coli* cultures. The obtained dodecane was diluted at 1:100 in hexane for GC-MS analysis. The GC/MS-TOF was equipped either with a VF-WAXms column (Agilent Technologies; 30 m x 0.25 mm i.d., 0.25 µm film thickness) or a DB-5ms column (Agilent Technologies; 30 m x 0.25 mm i.d., 0.25 µm film thickness), and the system was operated on the following conditions: (1) VF-WAXms, compounds were detected in split
10 mode at split ratio of 10:1, the GC oven temperature was programmed to ramp from 80 °C (held for 2 min) to 240 °C (held for 5 min) at 10 °C x min⁻¹; (2) DB-5ms, compounds were detected in split mode at split ratio of 10:1, the GC oven temperature was programmed to ramp from 50 °C (held for 2 min) to 160 °C at 10 °C x min⁻¹, to 230 °C at 8 °C x min⁻¹ and finally to 320 °C (held for 3 min). Mass spectra were acquired in the mass range of 33 300
15 m/z at the acquisition rate of 2 spectra/s. Ionization was performed by electron impact at 70 eV with an ion source temperature set at 230 °C.

[00127] *Structural Identification of Terpenoids*

[00128] Mass spectra obtained by electron ionization mode were used for initial compound identification by comparing them with the spectra of terpenoids in the National Institute of
20 Standards and Technology (NIST) database and published terpene spectra. Furthermore, Kovats retention indices of compounds produced were identified by calibrating with GC-MS with a C8-C30 alkane mix and were compared to the published retention indices in literature or in the NIST database. Major terpene products were verified, whenever possible, by comparison of retention time and mass spectra with authentic standards or essential oils with
25 known terpene compositions. Niaouli essential oil [viridiflorene 6 (10.1% w/w), viridiflorol 7 (18.1% w/w)], Cedrela woods oil [α -muurolene 5 (1% w/w), δ -cadinene 4 (11.7% w/w)], Cubeb oil [germacrene D (1% w/w), γ -muurolene 2 (4.2% w/w), β -cubebene (4.4% w/w), cubebol (15.2% w/w)], Amyris wood oil [β -elemene (germacrene A) (0.1% w/w), δ -cadinenol, 0.2%]. In addition, the structure of Δ^6 -protoilludene was further confirmed by
30 nuclear magnetic resonance spectroscopy.

[00129] *Functional Annotation for Terpene Synthases and its Gene Clusters in the *Agrocybe aegerita* Genome*

[00130] All the predicted amino acid sequences of protein-coding genes present in the genome of the dikaryotic strain *A. aegerita* AAE-3 have been searched for homologues to already characterized sesquiterpene synthases of *Coprinopsis cinerea*, *Omphalotus olearius* and *Stereum hirsutum* by blastp using Geneious® (version 9.1.8, Biomatters Ltd., Auckland, New Zealand). The predicted TPSs genes were then manually annotated. In addition, antiSMASH analysis was performed using the BiosynML plugin for Geneious® to predict terpene gene clusters in the *A. aegerita* genome.

[00131] *Cloning and Expression of Terpene Synthase Genes in E. coli*

[00132] Candidate fungal TPS genes were synthesized by Integrated DNA technologies and codon-optimized for expression in *E. coli*. The genes were cloned into pET11a vector for expression under the control of the T7 promoter. The resulting plasmid was transformed into *BL 21* strains carrying the plasmid p15A-cam-T7-dxs-idi which was redesigned from the plasmid pACM-T7-dxs-T7-idi-T7-ADS-ispA. Furthermore, the dxs in the plasmid was mutated to SL3 or SL5 (Fig. 2) to improve the solubility and activity. Single colony of the transformed *E. coli* cells was inoculated into 4 ml ZYM5052 auto-inducing medium (1% tryptone, 0.5% yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 0.5% glycerol, 0.05% glucose, 0.2% α -lactose) (Studier, 2005) with ampicillin (100 mg/L) and chloramphenicol (33 mg/L). After 14 h of cultivation at 28 °C and 250 rpm on a shaking incubator, the culture fluid was transferred into a 20 mL headspace screw top vials (Merck) and the headspace was sampled at 50 °C for 15 min by SPME.

[00133] *Homology Searches and Phylogenetic Tree Construction*

[00134] The 11 *A. aegerita* TPSs were used to search other fungal TPSs in Basidiomycota and Ascomycota genomes sequenced and published by the Joint Genome Institute under the Fungal Genomics Program (<http://genome.jgi-psf.org/programs/fungi/index.jsf>) and in the UniProt database by Basic Local Alignment Search Tool program (<http://www.uniprot.org/blast/>). In addition, the previously published 392 basidiomycota TPSs were incorporated. The combined TPS candidates were manually inspected for duplicate sequences, erroneous protein predictions, such as incomplete sequences that deviated from the expected protein length (200–800 aa, except for two putative TPSs, Disac1_349444 and EXIGL_831178) or lacking the conserved metal-binding DxxxD and NSE/DTE triad, or with predicted additional domains (such as geranylgeranyl pyrophosphate synthase functions). Upon identification of putative TPS amino acid sequences, their

alignments were performed using Clustal Omega and phylogenetic analyses were conducted with the Neighbor-Joining method using Clustal Omega or MEGA version 7.0.26.

[00135] *Analysis of TPS homologues by sequence similarity networks*

[00136] The curated fungal TPSs were analyzed by Enzyme Function Initiative-Enzyme
5 Similarity Tool (EFI-EST) web tool (<http://efi.igb.illinois.edu/efi-est/>) to generate sequence similarity networks (SSNs). The resulting SSNs were visualized using the open source software Cytoscape (<http://www.cytoscape.org/>). Inspection of the resulting SSNs is essential to obtain isofunctional clusters. Based on the SSNs generated by EFI-EST and sequentially varying a series of database-independent alignment score, a group of putative isofunctional
10 groups (PIGs) were obtained. The data of PIGs and traditional phylogenetic trees were compared to select the putative isofunctional TPSs. Here, the three novel TPSs (viridiflorol synthase AAE3_13291, viridiflorene synthase AAE3_12839, and linalool/nerolidol synthase AAE3_9435) were chosen to probe other putative isofunctional TPSs which were further validated by experiments.

15 [00137] *Δ 6-Protoilludene extraction and NMR validation of its structure*

[00138] The AAE3_10454 recombinant *E. coli* strain was cultured in 200 mL of ZYM5052 auto-inducing medium, supplemented with 100 mg of the spherical C18 resin (VersaFlash spherical C18 bonded flash silica 45-75 μ m, Sigma-Aldrich). After 24 h of cultivation at 28 °C and 150 rpm, the cell culture was manually filtered by a C18 cartridge and was
20 subsequently washed twice by deionized water. After filtration, the cells and liquid media were removed from the C18 cartridge. The terpene compound bound to the C18 resin was eluted by 10 mL of hexane. The eluted terpene solution was evaporated at 4 °C and subsequently analyzed on a Bruker DRX-400 NMR spectrometer with Cryoprobe, using 5-mm BBI (1H, G-COSY, multiplicity-edited G-HSQC, and G-HMBC spectra) or BBO (13C
25 spectra) probe heads equipped with z-gradients. Spectra were calibrated to residual protonated solvent signals CHCl₃ δ H 7.24 and CDCl₃ δ C 77.23). The terpene compound was verified as Δ 6-protoilludene by comparing the NMR spectral data with those reported in the literature.

30 [00139] **Example 1**

[00140] *Engineering an E. coli strain for TPS characterization.*

[00141] The wild-type *E. coli* BL21 produces little amount of terpene precursors (GPP and FPP), therefore, it is not suitable as a TPS characterization platform. To improve the detection sensitivity and accuracy, DXS and IDI were overexpressed to improve the intracellular precursors (Figure 3). Distinct from existing methods, two DXS mutants (SL3 and SL5) were identified based on random mutagenesis and screening. As shown in Figure 2, SL3 and SL5 had higher solubility over wild-type DXS, and therefore a higher activity than wild-type DXS. More importantly, SL5 has higher specific activity than wild-type DXS (Figure 2B). As a result, the lycopene yield in the strain overexpressing SL3 or SL5 was higher than that of wild-type DXS. Here, lycopene was used as an indicator to prove that GPP and FPP in the strains (SL3 and SL5) are higher. With the DXS mutants, the detection sensitivity of the cell platform is further improved. Hence, the *E. coli* strain (SL3 or SL5) was used as the platform for characterization of TPSs. Figure 27 shows a crystal structure of DXS where beneficial mutations have been highlighted. The mutants are related to the improved solubility of DXS by enzyme engineering approach. The ligand pyrophosphate was shown in salmon color and magnesium was in firebrick color.

[00142] **Example 2**

[00143] *Analysis of terpenes produced in A. aegerita*

[00144] To obtain an estimate of terpenes produced in *A. aegerita*, volatile compounds produced by its liquid cultures were analyzed. The illudin precursor, $\Delta(6)$ -protoilludene **1**, was a dominant metabolite produced by *A. aegerita* (Figure 1). In addition, small amount of α -ioscomene[#], α -, β -cubebene, β -copaene, γ -muurolene **2**, δ -cadinene **4**, β -selinene[#] **9**, cubenene, epicubenol **10** and cubenol (Figure 4 and 5 for chemical structure and mass spectra of all the terpene identified in the study, respectively) were observed after 26 days of culture. The results proved that the mushroom, *A. aegerita* produces structurally diverse terpenes.

[00145] **Example 3**

[00146] *The sesquiterpenome of A. aegerita*

[00147] During fructification of *A. aegerita* 20 putative terpenoids were detected by means of GC/MS analysis, of which the tentatively identified $\Delta 6$ -protoilludene[#] was the most prominent compound (Figure 1). Other major compounds were α -cubebene, α -isocomene[#], β -cubebene and δ -cadinene (Compound structures in Figure 4, the mass spectra in Figure 5).

The blastp search for putative STSs present in the genome of *A. aegerita* revealed 11 genes (Table 2 and Figure 6). Seven of the TPSs cluster with already known basidiomycete TPSs into at least three different groups. Four putative TPSs (AAE3_09008, AAE3_06743, AAE3_04444 and AAE3_05024) compile to an own cluster (Figure 6). Four of the STS genes (AAE3_10454, AAE3_12839, AAE3_04120 and AAE3_13291) are part of clusters consists amongst other of two to five P450 monooxygenases.

[00148] **Table 2.** Details on *Agrocybe aegerita* STSs genes.

Protein ID	scaffold	gene start	gene stop	gene length	number of introns	protein length
04120	2	9,526	11,372	1,847	6	659
04120 short	2	10,043	11,372	1,330	5	346
04444	2	1,033,830	1,035,120	1,291	4	353
09164	4	405,253	406,500	1,248	4	342
13190	8	106,456	107,896	1,441	6	358
13291	8	437,487	439,057	1,571	5	430
05024	21	111,488	112,812	1,325	4	355
06595	28	328,403	329,611	1,209	3	346
06743	29	231,813	233,188	1,376	4	372
09008	39	347,841	349,082	1,242	6	308
10454	49	17,315	18,741	1,427	5	387
12839	70	55,035	56,437	1,403	4	389

[00149] **Example 4**

10 [00150] *Characterization of 11 Predicted Sesquiterpene Synthases*

[00151] All 11 predicted STSs were codon optimized and cloned into the pET vector, which was transformed into an engineered *E.coli BL21* strain overproducing farnesyl pyrophosphate (FPP), the sesquiterpene precursor. Compounds tentatively identified on basis of their retention index (RI) and mass spectra in comparison to those in the literature and
15 databases as described in the methods are marked with a hashmark (#).

[00152] All STSs (TPSs) except AAE3_09008 and AAE3_05024 gave rise to one or more sesquiterpenes in liquid cultures of the corresponding *E. coli* clone (Figure 7 (DB-5ms column), Figure 8 (VFWAXms column) and Table 3). AAE3_04120 and AAE3_10454 produced the same sesquiterpene as the only product. NIST database search and a comparison
20 with mass spectra of fungal sesquiterpenes from previous reports (Figure 5 and Table 3) revealed this compound could be Δ^6 -protoilludene **1**. And its structure was further validated by the NMR analysis which had the identical spectrum with previous report (Figure 9). Δ^6 -

protoilludene is the precursor for illudins that have shown anti-tumor and antimicrobial effects. Till now, six Δ^6 -protoilludene synthases from three fungal species have been reported, Omp6 and Omp7 from *Omphalotus olearius*, Pro1 from *Armillaria gallica* and Steh1_25180, Steh1_64702 and Steh1_73029 from *Stereum hirsutum*. Interestingly,

5 AAE3_04120 and AAE3_10454 form a closely related subgroup with the six reported synthases in the phylogenetic clustering (Figure 1), indicating that the 6-protoilludene synthases are highly related among different fungal species.

[00153] **Table 3.** Terpene products of the TPSs in this study.

SPME data		DB5 (Non-polar)			VFWAXms (Polar)		
Gene	Products	Area %	RI	Literature RI	Area %	RI	Literature RI
AAE3_04120	Δ^6 -protoilludene	100%	1391	1393	100%	1513	/
AAE3_04444	β -elemene	8%	1400	1391 $\pm$ 2 (521)	/	/	1591 $\pm$ 9 (259)
	γ -muurolene	33%	1487	1477 $\pm$ 3 (392)	30%	1706	1692 $\pm$ 12 (165)
	β -selinene	18%	1509	1486 $\pm$ 3 (349)	/	/	1717 $\pm$ 13 (167)
	α -selinene	14%	1515	1517	/	/	1656 $\pm$ 0 (2)
	β -cadinene	21%	1518	1518 $\pm$ 10 (30)	22%	1733	1720 $\pm$ N/A (1)
	δ -cadinene	6%	1529	1524 $\pm$ 2 (751)	4%	1772	1758 $\pm$ 13 (374)
	α -epi-Cadinol	/	/	/	9%	2213	2169 $\pm$ 16 (145)
AAE3_06595	γ -muurolene	5%	1489	1477 $\pm$ 3 (392)	/	/	1692 $\pm$ 12 (165)
	β -selinene	5%	1508	1486 $\pm$ 3 (349)	/	/	1717 $\pm$ 13 (167)
	α -muurolene	7%	1511	1499 $\pm$ 3 (427)	8%	1739	1726 $\pm$ 13 (198)
	α -Selinene	5%	1514	1517	/	/	1656 $\pm$ 0 (2)
	δ -cadinene	60%	1529	1524 $\pm$ 2 (751)	57%	1772	1758 $\pm$ 13 (374)
	T-muuroloI	8%	1659	/	13%	2198	2186 $\pm$ 16 (140)
	α -cadinol	10%	1671	/	18%	2243	2226 $\pm$ 9 (182)
AAE3_6743	γ -muurolene	27%	1488	1477 $\pm$ 3 (392)	14%	1706	1692 $\pm$ 12 (165)
	β -cadinene	13%	1516	1518 $\pm$ 10 (30)	13%	1733	1720 $\pm$ N/A (1)
	δ -cadinene	13%	1527	1524 $\pm$ 2 (751)	6%	1772	1758 $\pm$ 13 (374)
	Unknown sesquiterpene alcohol	44%	1650	/	52%	2176	
AAE3_9164	β -myrcene*	10%	989	991 $\pm$ 2 (841)	7%	1172	1161 $\pm$ 7 (569)
	β -copaene	/	/	/	2%	1610	1586 $\pm$ 12 (15)
	α -cubebene	8%	1358	1351 $\pm$ 2 (338)	1%	1469	1463 $\pm$ 6 (186)
	γ -muurolene	15%	1487	1477 $\pm$ 3 (392)	8%	1677	1692 $\pm$ 12 (165)
	δ -cadinene	37%	1531	1524 $\pm$ 2 (751)	59%	1772	1758 $\pm$ 13 (374)
	epizonarene	13%	1538	1501 $\pm$ 4 (28)	8%	1677	1677 $\pm$ 1 (15)
	Unknown sesquiterpene	9%	1548	/	4%	1736	1786 $\pm$ 13 (65)
	epicubenol	8%	1660	1627 $\pm$ 2 (144)	7%	2076	2067 $\pm$ 21 (67)
	cubebol	/	/	/	7%	1951	1957
AAE3_10454	Δ^6 -protoilludene	100%	1392	0	100%	1513	/

AAE3_12839	δ -elemene	2%	1345	1338 $\pm$ 2 (221)	/	/	1470 $\pm$ 9 (86)
	(+)-aromadendrene	7%	1457	1440 $\pm$ 1 (5)	/	/	1635 $\pm$ 2 (3)
	viridiflorene	82%	1508	1493 $\pm$ 4 (114)	81%	1712	1697 $\pm$ 7 (76)
	Unknown sesquiterpene	9%	1514	/	/	/	/
AAE3_13190	γ -muurolene	22%	1489	1477 $\pm$ 3 (392)	9%	1706	1692 $\pm$ 12 (165)
	(-)-germacrene D	/	/	/	4%	1730	1710 $\pm$ 14 (325)
	α -muurolene	32%	1512	1499 $\pm$ 3 (427)	20%	1740	1726 $\pm$ 13 (198)
	δ -cadinene	24%	1529	1524 $\pm$ 2 (751)	24%	1772	1758 $\pm$ 13 (374)
	Cubenol	/	/	/	4%	2076	2080 $\pm$ 4 (65)
	δ -cadinol/ δ -cedrol	21%	1662	1645	24%	2209	2187 $\pm$ 10
AAE3_13291	viridiflorene	42%	1509	1493 $\pm$ 4 (114)	9%	1714	1697 $\pm$ 7 (76)
	viridiflorol*	58%	1617	1591 $\pm$ 2 (198)	91%	2099	2095 $\pm$ 10 (108)
Denbi1_816208	viridiflorene	31%	1509	1493 $\pm$ 4 (114)	8%	1714	1697 $\pm$ 7 (76)
	viridiflorol*	57%	1617	1591 $\pm$ 2 (198)	92%	2099	2095 $\pm$ 10 (108)
Sphst_47084	viridiflorene	32%	1509	1493 $\pm$ 4 (114)	8%	1714	1697 $\pm$ 7 (76)
	viridiflorol*	58%	1617	1591 $\pm$ 2 (198)	92%	2099	2095 $\pm$ 10 (108)
Pilcr_825684	β -elemene	7%	1401	1391 $\pm$ 2 (521)	/	/	1591 $\pm$ 9 (259)
	Unknown sesquiterpene	12%	1490	/	/	/	/
	Viridiflorene	6%	1509	1493 $\pm$ 4 (114)	8%	1714	1697 $\pm$ 7 (76)
	epi- α -Selinene	8%	1514	1485 $\pm$ N/A (1)	12%	1726	1725
	γ -cadinene	45%	1529	1513 $\pm$ 2 (485)	30%	1774	1765 $\pm$ 11 (199)
	Unknown sesquiterpene	15%	1538	/	/	/	/
Galma_104215	β -gurjunene	83%	1441	1432 $\pm$ 3 (234)	81%	1673	1669 $\pm$ 17 (14)
	Unknown sesquiterpene	10%	1439	/	12%	1634	/

- [00154] The *E. coli* strains expressing AAE3_0444 (SEQ ID NO: 27) and AAE3_6743 (SEQ ID NO: 29) produced several sesquiterpene compounds (Figure 7, Figure 8 and Table 3). γ -Muurolene **2** and β -cadinene[#] **3** are the main products of AAE3_0444, together accounting for >50% of the total sesquiterpenes detected. In addition, small amounts of β -selinene[#] **9**, α -selinene[#] **13**, β -elemene (germacrene A) **11** and δ -cadinene **4** (verified by Cube essential oil, Figure 10 and S7) were detected in the headspace of AAE3_0444 culture. AAE3_6743 produced an unknown sesquiterpenol as the main product, together with small amounts of γ -muurolene **2** (27%), β -cadinene[#] **3** (13%) and δ -cadinene **4** (13%).
- [00155] A wide variety of sesquiterpenes were detected for the *E. coli* culture expressing AAE3_09164 (SEQ ID NO: 30) (Figure 7, Figure 8 and Table 3). Among them, δ -cadinene **4** (37%) was main product, together with γ -muurolene **2** (15%), epizonarene[#] (13%),

epicubenol[#] **10** (8%), α -cubebene[#] (8%) and many uncharacterized minor products. In addition, cubebol (6.7%, Table 3) was detected by VFWAXms (but not by DB5 column) and verified by Cubeb essential oil (Figure 10). Interesting, a noticeable amount of the monoterpene β -myrcene (10%, verified by authentic standard, Figure 12) was detected in the headspace of AAE3_09164 culture, despite that the *E. coli* strain produced only little amount of the monoterpene precursor geranyl pyrophosphate (GPP). The results suggested AAE3_09164 could be a bi-functional enzyme that is able to use both FPP and GPP as substrates to synthesize sesquiterpenes and monoterpenes, respectively. Similar bi-functional TPSs were reported previously in the ascomycete family Hypoxylaceae, such as Hyp4, Hyp5 from *Hypoxylon sp.* and EC12-PGS from *Daldinia eschscholzii*. In the phylogenetic analysis of the deduced AAE3_09164 amino acid sequence clustered together with Cop4, Omp4 and Steh1_128017 enzymes (Figure 13). Indeed, all of these enzymes including AAE3_09164 are highly promiscuous enzymes with δ -cadinene **4** as a common major product.

[00156] The *E. coli* strain expressing AAE3_13190 (SEQ ID NO: 33) produced four major products, α -muurolene **5** (32%) and γ -muurolene **2** (22%), δ -cadinene **4** (24%) and δ -cadinol[#] **8** (21%) (Figure 7, 8, 10 and 11 and Table 3). In addition, there were at least six other minor sesquiterpene products, including (-)-germacrene D (Table 3) and verified by verified by Cubeb essential oil (Figure 10). According to phylogenetic clustering in Figure 13, AAE3_13190 is closely related to Cop3 from *Coprinopsis cinerea* and Omp3 from *Omphalotus olearius*. Consistently, all of them produced α -muurolene **5** as the major product. The major product for the *E. coli* culture expressing AAE3_06595 (SEQ ID NO: 28) was δ -cadinene **4** (60% of total terpenes). In addition, a few minor sesquiterpene compounds were also detected for AAE3_06595 culture including γ -muurolene **2**, β -selinene[#] and T-muurolol[#]. The enzymes Cop1 and Omp2 are closely related to AAE3_06595. However, Omp2 was not functional in *E. coli*, and δ -cadinene **4** was only one minor product of Cop1 whose main product was β -elemene.

[00157] The major product of AAE3_12839 (SEQ ID NO: 34) was viridiflorene **6**. In contrast, the *E. coli* strain expressing AAE3_13291 (SEQ ID NO: 32) produced viridiflorol **7** as the major product (viridiflorol **7** and viridiflorene **6** were confirmed by Niaouli essential oil, Figure 14), with small amount of viridiflorene **6** (8.6% with VFWAXms in Table 3 and Figure 8). Here, the data of VFWAXms column instead of DB5 column was used to quantify viridiflorol **7**, as the quantification of viridiflorol **7** in DB5 was inaccurate with a significantly

lower signal than viridiflorene 6 (Figure 7 and Table 3). To our knowledge, no viridiflorene synthase or viridiflorol synthase has been reported in fungi. Even in plants, only six viridiflorene synthases were identified from *Solanum lycopersicum* and *Nicotiana tabacum* (Common tobacco). The alignment of AAE3_12839 and the tomato viridiflorene synthase indicated that there was limited sequence similarity (Figure 15). Similarly, AAE3_13291 shares only 11% identity and less than 30% similarity with the viridiflorol synthase from *Melaleuca quinquenervia*, which is the only viridiflorol synthase reported so far (Figure 16). [00158] Furthermore, the identified TPSs in *A. aegerita* shared the same first cyclization step with TPSs in *C. cinerea* and *O. olearius*. For Δ -6-protoilludene synthase (AAE3_4120 (SEQ ID NO: 26) and AAE3_10454 (SEQ ID NO: 31)), viridiflorene synthase (AAE3_12839 (SEQ ID NO: 34)) and viridiflorol synthase (AAE3_13291 (SEQ ID NO: 32)), they all proceed through 1,11 cyclization of FPP to form tricyclic sesquiterpenes (Figure 17). In contrast, other TPSs (AAE3_04444 (SEQ ID NO: 27), AAE3_6595 (SEQ ID NO: 28), AAE3_9164 (SEQ ID NO: 30), AAE3_13190 (SEQ ID NO: 33) and AAE3_6743 (SEQ ID NO: 29)) preferentially catalyze a 1,10 cyclization of FPP to form bicyclic sesquiterpenes.

[00159] **Example 5**

[00160] *Analysis of fungal genome for TPS functional study*

[00161] The results in Figure 7 and 13 reinforced that certain types of fungal TPSs have highly conserved sequences fortified by identical products, such as eight characterized Δ -6-protoilludene synthases and four characterized δ -cadinene synthases. Thus, phylogenetic analysis provides a predictive framework to identify novel terpene synthases with novel or similar functions. The predictive accuracy of the model increases as the number of experimentally characterized TPSs accumulates. Previously, three of the Δ -6-protoilludene synthases (Stehi1_25180, Stehi1_64702 and Stehi1_73029 from *S. hirsutum*) were correctly predicted and validated through bioinformatic analysis. Since then, the genomes of many new fungal species have been sequenced but their TPS genes have not been studied. Here, the aim was to establish a new predictive framework for the functional study of uncharacterized fungal TPSs with the new characterized *A. aegerita* TPSs and previously studied fungal TPSs. Through BLAST search in fungal genome database at the Joint Genome Institute (JGI, <http://genome.jgi-psf.org/programs/fungi/index.jsf>) and in The Universal Protein Resource (UniProt, <http://www.uniprot.org/>), about 2,000 putative TPS genes was uncovered. After a

series of curation (as described in methods), a total of 1,408 putative TPSs from 85 Basidiomycota and 239 Ascomycota genomes were obtained (Table 4). On average, Basidiomycota have an average of 10-15 TPSs per genome (800 TPSs from 84 Basidiomycota) but about 80% Ascomycota have only 1-3 TPSs per genome (594 TPSs from 5 236 Ascomycota).

[00162] **Table 4.** The information about 1408 putative fungal TPSs in this study.

Ascomycota	Basidiomycota
Aaosphaeria arxii	Agaricus bisporus
Aaoar1_459904	Agabi_varbisH97_2_119105
Acephala macrosclerotiorum	Agabi_varbisH97_2_144791
Aciaci1_473652	Agabi_varbisH97_2_149463
Acremonium strictum	Agabi_varbisH97_2_195544
Alternaria alternata	Agabi_varbisH97_2_73543
Alta11_1080498	Agabi_varbur_1_109605
Alternaria brassicicola	Agabi_varbur_1_126555
Altbr1_7288	Agabi_varbur_1_130532
Amniculicola lignicola	Agabi_varbur_1_46681
Amnli1_450732	Agabi_varbur_1_61902
Amore1_23054	Agabi_varbur_1_76352
Ampelomyces quisqualis	Agabi_varbur_1_79290
Ampqui1_550807	Agrocybe aegerita
Anthostoma avocetta	AAE3_04120
Antav1_377590	AAE3_04444
Antav1_383196	AAE3_05024
Antav1_400494	AAE3_06595
Antav1_445568	AAE3_06743
Antav1_446501	AAE3_09008
Antav1_453578	AAE3_09164
Antav1_468055	AAE3_10454
Antav1_472246	AAE3_109435
Antav1_476690	AAE3_12839
Antav1_484797	AAE3_13190
Antav1_504933	AAE3_13291
Apiospora montagnei	Agrocybe pediades
Apimo1_107765	Agrped1_109003
Apimo1_109481	Agrped1_640059
Aplosporella prunicola	Agrped1_665597
Aplpr1_315168	Agrped1_689671
Arthrotrrys oligospora	Agrped1_689675
Artol1_6616	Agrped1_693394
Arthroderma benhamiae	Agrped1_705454
Artbe1_2427	Agrped1_749682

Ascocoryne sarcoides	Agrped1_804989
Ascsa1_1273	Agrped1_804996
Ascsa1_6084	Agrped1_820868
Aspergillus aculeatinus	Amanita muscaria
Aspacu1_414218	M378_161967
Aspacu1_433825	M378_167361
Aspergillus brasiliensis	M378_181109
Aspbr1_199648	M378_186936
Aspergillus brunneoviolaceus	M378_457656
Aspbru1_469179	M378_74452
Aspergillus calidoustus	M378_78547
Aspcal1_764165	M378_9904
Aspcal1_767797	Armillaria gallica
Aspcal1_768162	Pro1
Aspcam1_281412	Auricularia delicata
Aspcam1_337372	Aurde1_106904
Aspergillus carbonarius	Aurde1_129583
Aspca3_517619	Aurde1_138561
Aspcl1_4114	Aurde1_166047
Aspergillus costaricaensis	Aurde1_173663
Aspcos1_212514	Aurde1_56959
Aspcos1_272862	Aurde1_61813
Aspergillus fijiensis	Aurde1_62781
Aspfij1_393093	Aurde1_73423
Aspergillus flavus	Aurde1_73447
Aspfl1_36410	Aurde1_73578
Aspergillus heteromorphus	Aurde1_75612
Asphet1_431105	Aurde1_81767
Aspergillus homomorphus	Aurde1_90621
Asphom1_411924	Aurde1_97553
Aspergillus ibericus	Auricularia subglabra
Aspibe1_454210	AURDE_130623
Aspergillus indologenus	Bjerkandera adusta
Aspind1_388535	Bjead1_1_105488
Aspergillus kawachii	Bjead1_1_117829
Aspka1_1_17804	Bjead1_1_156307
Aspka1_1_20838	Bjead1_1_158616
Aspergillus lacticoffeatus	Bjead1_1_166045
Asplac1_345547	Bjead1_1_172777
Asplac1_444313	Bjead1_1_337295
Aspergillus luchuensis	Bjead1_1_53082
Aspfo1_40412	Bjead1_1_54261
Aspfo1_48364	Bjead1_1_54262
Aspfo1_701161	Bjead1_1_64972

Aspergillus neoniger	Botryobasidium botryosum
Aspneo1_451579	Botbo1_115253
Aspergillus niger	Botbo1_147563
Aspni_byT_1_291648	Botbo1_150401
Aspni_byT_1_339193	Botbo1_177898
Aspni_DSM_1_158481	Botbo1_189629
Aspni_DSM_1_165991	Botbo1_35044
Aspni_NRR3_1_492	Calocera cornea
Aspni_NRR3_1_8436	CALCO_485200
Aspni_NRR3_1_8732	Calocera viscosa
Aspni7_1085752	CALVI_546272
Aspni7_1155978	CALVI_549316
Aspergillus nomius	CALVI_565570
Aspnom13137_1_4577	Ceriporiopsis subvermisporea
Aspnom13137_1_5237	Cersu1_100300
Aspnom13137_1_5921	Cersu1_107906
Aspergillus novofumigatus	Cersu1_108146
Aspoch1432_1_2847	Cersu1_113927
Aspergillus oryzae	Cersu1_114263
Aspor1_10090	Cersu1_116249
Aspergillus phoenicis	Cersu1_126560
Asppl1_338445	Cersu1_161387
Aspergillus piperis	Cersu1_162846
Asppip1_454731	Cersu1_162851
Aspergillus sclerotii carbonarius	Cersu1_52233
Aspscle1_371398	Cersu1_71514
Aspergillus steynii	Cersu1_78286
Aspste1_453294	Cersu1_83362
Aspergillus terreus	Cersu1_85360
Aspte1_5331	Cersu1_95867
Aspergillus udagawae	Cersu1_96486
Aspuda1_1612	Cersu1_98094
Aspuda1_4266	ter14
Aspergillus vadensis	Coniophora puteana
Aspvad1_340387	Conpu1_102165
Aspvad1_341847	Conpu1_102220
Aspwe1_186729	Conpu1_118913
Aspwe1_691717	Conpu1_137465
Aureobasidium pullulans	Conpu1_152083
Aurpu_var_mell_89219	Conpu1_155138
Baudoinia compniacensis	Conpu1_156845
Bauco1_152112	Conpu1_15871
Bimuria novae-zelandiae	Conpu1_168606
Biscogniauxia nummularia	Conpu1_170276

Bisnum1_472611	Conpu1_47697
Bisnum1_480590	Conpu1_50941
Bisnum1_560481	Conpu1_58009
Bisnum1_595288	Conpu1_58901
Bisnum1_611126	Conpu1_58994
Bisporella sp.	Conpu1_60451
Bisspl_639301	Conpu1_62719
Bisspl_741721	Conpu1_62911
BcBOT2	Conpu1_63003
Bysci1_371003	Conpu1_75631
Cadophora sp.	Conpu1_88505
Cadsp1_422591	Conpu1_92191
Caloscypha fulgens	Coprinopsis cinerea
Calful1_769187	CC1G_03587
Capronia epimyces	Cop1
Capep1_3727	Cop2
Chaetomium globosum	Cop3
CHGG_03509	Cop4
Chalara longipes	Cop5
Chalo1_381634	Cop6
Chalo1_464358	Cylindrobasidium torrendii
Cladophialophora bantiana	CYLTO_347245
Claba1_132379	CYLTO_369585
Cladophialophora psammophila	CYLTO_380537
Claps1_13034	CYLTO_384541
Cladorrhinum bulbillosum	CYLTO_400743
Clabul1_1016528	CYLTO_405471
Clabul1_76434	CYLTO_436484
Clabul1_847239	CYLTO_442632
Clathrospora elyae	CYLTO_452977
Clael1_510577	CYLTO_453006
Coccomyces strobili	Dacryopinax primogenitus
Cocst1_631366	DACRY_34691
CocheC4_1_36610	Dacryopinax sp.
CocheC5_3_10970	Dacsp1_109687
Cochliobolus miyabeanus	Dacsp1_81212
Cocmi1_93348	Dacsp1_96371
Cochliobolus sativus	Daedalea quercina
Cocsa1_348577	DAEQU_261749
Colac2_589620	DAEQU_662879
Colac2_693029	DAEQU_663038
Colac2_720284	DAEQU_677968
Colac2_722687	DAEQU_696090
Colac2_756572	DAEQU_737681

Colletotrichum caudatum	DAEQU_745062
Colca1_582509	DAEQU_769721
Colca1_613400	DAEQU_811112
Colletotrichum cereale	Dendrothele bispora
Colce1_637756	Denbi1_650172
Colce1_710743	Denbi1_654460
Colce1_751683	Denbi1_659367
Colce1_753190	Denbi1_667929
Colletotrichum eremochloae	Denbi1_678334
Coler1_553160	Denbi1_689487
Coler1_633162	Denbi1_690253
Coler1_645427	Denbi1_692356
Colletotrichum fioriniae	Denbi1_693874
Colfi1_276541	Denbi1_750040
Colfi1_276864	Denbi1_792287
Colfi1_283382	Denbi1_816208
Colfi1_285486	Denbi1_818935
Colfi1_288712	Denbi1_824130
Colletotrichum godetiae	Denbi1_855029
Colgo1_546119	Denbi1_866377
Colgo1_562331	Denbi1_873510
Colgo1_645279	Denbi1_896419
Colgo1_696718	Diaporthe helianthi
Colgo1_730749	DHEL01_07884
Colletotrichum higginsianum	Dichomitus squalens
Colhig2_12235	Dicsq1_104353
Colhig2_13496	Dicsq1_138476
Colhig2_6613	Dicsq1_144469
Colhig2_7207	Dicsq1_146430
Colhig2_9460	Dicsq1_147637
Collu1_212508	Dicsq1_159719
Collu1_590124	Dicsq1_170641
Collu1_79349	Dicsq1_181048
Colletotrichum navitas	Dicsq1_57723
Colna1_600097	Dicsq1_58025
Colna1_637650	Dicsq1_63165
Colny1_1016018	Dicsq1_80177
Colny1_1018170	Dicsq1_80370
Colny1_1022050	Dicsq1_86568
Colny1_1022440	Exidia glandulosa
Colletotrichum orchidophilum	EXIGL_605329
Color1_5151	EXIGL_611671
Color1_6973	EXIGL_620059
Color1_848	EXIGL_664938

Colletotrichum phormii	EXIGL_673075
Colph1_306140	EXIGL_673208
Colph1_417792	EXIGL_677911
Colph1_464784	EXIGL_677941
Colph1_466218	EXIGL_680198
Colph1_479875	EXIGL_681577
Colph1_516153	EXIGL_688085
Colsa1_939591	EXIGL_713320
Colsa1_940033	EXIGL_743228
Colsa1_941201	EXIGL_750528
Colsa1_942596	EXIGL_767126
Colsa1_948955	EXIGL_769607
Colsa1_950600	EXIGL_769609
Colletotrichum simmondsii	EXIGL_770624
Colsi1_971930	EXIGL_773846
Colsi1_972523	EXIGL_831178
Colsi1_972624	Fibroporia radiculosa
Colsi1_976172	FIBRA_00633
Colsi1_976953	FIBRA_00800
Colsi1_979039	FIBRA_05385
Colsi1_981054	FIBRA_05798
Colsi1_981282	FIBRA_06228
Colsi1_983009	FIBRA_06230
Colso1_559351	FIBRA_06895
Colletotrichum sublineola	FIBRA_07171
Colsu1_648985	FIBRA_07173
Colsu1_724576	Fibulorhizoctonia sp.
Colzo1_706815	FIBSP_768030
Coniella sp	FIBSP_820394
Pilidi1_186199	FIBSP_832548
Coniochaeta ligniaria	FIBSP_943511
Conli1_10674	Fistulina hepatica
Conli1_1914	FISHE_34696
Conlig1_583628	FISHE_45426
Conlig1_658201	FISHE_46267
Coniochaeta sp.	FISHE_66255
ConPMI546_932510	Fomitiporia mediterranea
ConPMI546_934988	Fomme1_105378
Coniosporium apollinis	Fomme1_109318
Conap1_98915	Fomme1_112446
Corollospora maritima	Fomme1_170128
Corma2_707499	Fomme1_17224
Cryphonectria parasitica	Fomme1_27083
Crypa2_343514	Fomme1_80051

Cryptodiaporthe populea	Fomme1_80204
Cryo1_327771	Fomme1_80444
Cryo1_328559	Fomme1_82079
Cryo1_335598	Fomme1_82792
Cryo1_345542	Fomme1_82811
Cryo1_376330	Fomme1_89798
Cryo1_381328	Fomme1_91806
Cryo1_381563	Fomme1_95393
Cryo1_432491	Fomme1_97061
Cryo1_443797	Fomitopsis pinicola
Cryo1_472123	Fompi3_1017321
Cucurbitaria berberidis	Fompi3_1017322
Cucbe1_280026	Fompi3_1023716
Daldinia eschscholzii	Fompi3_1034271
DalEC12_1_12539	Fompi3_110513
DalEC12_1_17536	Fompi3_1118553
DalEC12_1_24646	Fompi3_1118777
DalEC12_1_24764	Fompi3_1120393
DalEC12_1_25458	Fompi3_1137037
DalEC12_1_70183	Fompi3_88169
Decorospora gaudefroyi	Galerina marginata
Decga1_179458	Galma_104215
Delphinella strobiligena	Galma_1278404
Delst1_202989	Galma_1352301
Delst1_230429	Galma_137032
Delst1_365307	Galma_143861
Diaporthe ampelina	Galma_222029
Diaam1_7440	Galma_223690
Diaam1_7814	Galma_225678
Diaam1_8586	Galma_229201
Didymella zeae-maydis	Galma_245845
Didma1_13214	Galma_266794
Didymocrea sadasivanii	Galma_62552
Didsa1_432338	Galma_63553
Didsa1_459411	Galma_63556
Diplodia seriata	Galma_72334
Dipse1_2018	Galma_72397
Dissoconium aciculare	Galma_78470
Disac1_349444	Ganoderma sp.
Dothidotthia symphoricarpi	Gansp1_106195
Dotsy1_400389	Gansp1_115598
Endocarpon pusillum	Gansp1_116882
EndpusZ1_8494	Gansp1_118798
EndpusZ1_8851	Gansp1_119170

Entoleuca mammata	Gansp1_126698
Entma1_245693	Gansp1_143866
Entma1_278690	Gansp1_147418
Entma1_396117	Gansp1_151250
Entma1_410097	Gansp1_151266
Eutypa lata	Gansp1_151299
Eutla1_2536	Gansp1_155853
Eutla1_3565	Gansp1_164758
Eutla1_5251	Gansp1_166943
Exophiala aquamarina	Gansp1_41036
Exoaq1_8751	Gansp1_57109
Fonsecaea pedrosoi	Gansp1_57679
Fonpe1_8054	Gansp1_58158
Fusarium fujikuroi	Gansp1_58881
Ffsc4	Gansp1_81688
Ffsc6	Gansp1_85736
Fusfu1_1126	Gloeophyllum trabeum
Fusfu1_11322	Glotr1_1_103889
Fusfu1_14268	Glotr1_1_116237
Fusfu1_2062	Glotr1_1_117331
Fusfu1_6471	Glotr1_1_131990
STC3	Glotr1_1_47645
STC5	Glotr1_1_48290
Fusarium graminearum	Glotr1_1_64172
CLM1	Glotr1_1_78472
Fusgr1_10122	Glotr1_1_79917
Fusgr1_13217	Glotr1_1_80390
Fusgr1_2052	Grifola frondosa
Fusgr1_4586	COP3_0_A0H81_12697
Fusgr1_548	COP3_1_A0H81_08013
Fusarium oxysporum	COP3_2_A0H81_10954
Fusox2_10433	COP3_5_A0H81_08017
Fusox2_10434	COP4_0_A0H81_07725
Fusox2_10435	COP4_1_A0H81_07728
Fusox2_10673	Gymnopus luxurians
Fusox2_10675	Gymlu1_1012408
Fusox2_8564	Gymlu1_1024248
Fusarium sporotrichioides	Gymlu1_152409
FsTDS	Gymlu1_164402
Fusarium verticillioides	Gymlu1_179557
Fusve2_12377	Gymlu1_181084
Fusve2_1423	Gymlu1_239618
Fusve2_19	Gymlu1_240529
Fusve2_20	Gymlu1_242070

Fusve2_8588	Gymlu1_249731
Fusve2_8699	Gymlu1_249732
Glomerella acutata	Gymlu1_257858
Gloac1_1349405	Gymlu1_266288
Gloac1_1383433	Gymlu1_474275
Gloac1_1413417	Gymlu1_70394
Gloac1_1624359	Gymlu1_74039
Gloac1_1638878	Gymlu1_775187
Glomerella cingulata	Hebeloma cylindrosporum
Gloci1_1722638	M413_27416
Gloci1_1750922	M413_32803
Gloci1_1755285	M413_415200
Gloci1_1819074	M413_443011
Gloci1_1825757	M413_7659
Gloci1_1830608	M413_83524
Gloci1_1835014	Heterobasidion annosum
Gloci1_1852737	Hetan2_115814
Glonium stellatum	Hetan2_148791
Glost2_424907	Hetan2_167573
Gremmeniella abietina	Hetan2_169607
Greab1_510385	Hetan2_172256
Greab1_510929	Hetan2_181194
Grosmanina clavigera	Hetan2_34201
CMQ_352	Hetan2_382802
Groc1_2976	Hetan2_382866
Groc1_8310	Hetan2_42859
Gymnascella aurantiaca	Hetan2_446121
Gymau1_124723	Hetan2_454193
Gymau1_163306	Hetan2_458479
Gymnascella citrina	Hetan2_48772
Gymci1_1_287288	Hetan2_51706
Gyromitra esculenta	Hydnomerulius pinastri
Gyresc1_452646	HYDPI_175348
Gyresc1_614921	HYDPI_90513
Hirsutella minnesotensis	HYDPI_93448
HIM_03781	HYDPI_95823
Hymenoscyphus varicosporoides	Hypholoma sublateralitium
Hymvar1_186372	HYPsu1_151315
Hymvar1_433677	Hypsu1_138166
Hymvar1_527573	Hypsu1_138665
Hymvar1_530070	Hypsu1_148365
Hymvar1_530714	Hypsu1_148385
Hypoxyton sp.	Hypsu1_159396
Hyp1	Hypsu1_202683

Hyp2	Hypsu1_205915
Hyp3	Hypsu1_36467
Hyp4	Hypsu1_47068
Hyp5	Hypsu1_80866
HypCI4A_1_20984	Hypsu1_92421
HypCI4A_1_216497	Hypsizygus marmoreus
HypCI4A_1_322581	COP3_1_Hypma_09878
HypCI4A_1_59230	COP3_2_Hypma_09820
HypCI4A_1_6706	COP4_Hypma_01074
HypCI4A_1_69724	Jaapia argillacea
HypCI4A_1_7067	Jaaar1_125196
HypCO275_1_269219	Jaaar1_129042
HypCO275_1_31178	Jaaar1_162104
HypCO275_1_392541	Jaaar1_191378
HypCO275_1_397991	Jaaar1_192672
HypEC38_3_102477	Jaaar1_206626
HypEC38_3_372695	Jaaar1_35337
HypEC38_3_409185	Jaaar1_453389
HypEC38_3_424147	Jaaar1_47108
HypEC38_3_436214	Jaaar1_487951
Ilyonectria robusta	Jaaar1_62046
Ilyrob1_438077	Laccaria amethystina
Ilyrob1_458205	K443_108732
Ilyrob1_462532	K443_126876
Ilyonectria sp.	K443_309839
Ilysp1_1486196	K443_619353
Ilysp1_1873426	K443_681798
Kalaharituber pfeilii	K443_99583
Kalpfe1_784829	Laccaria bicolor
Kalpfe1_789340	LACBI_312850
Karstenula rhodostoma	LACBI_326872
Karrh1_427857	Lacbi1_297082
Karrh1_478359	Lacbi1_307420
Khuskia oryzae	Lacbi1_307559
Khuory1_125966	Lacbi1_307631
Khuory1_156064	Lacbi1_308775
Khuory1_357319	Lacbi1_310816
Khuory1_456225	Lacbi1_327169
Khuory1_483548	Lacbi1_331339
Khuory1_495123	Lacbi1_333748
Lecythophora sp.	Laetiporus sulphureus
LecAK0013_1_225655	LAESU_64487
LecAK0013_1_337743	LAESU_657286
LecAK0013_1_358472	LAESU_657700

Lentithecium fluviatile	LAESU_682207
Lenfl1_319520	LAESU_706375
Leptodontium sp.	LAESU_724692
Leptod1_444196	LAESU_736295
Leptod1_455689	LAESU_739029
Leptod1_476038	LAESU_754774
Leptosphaeria maculans	LAESU_760769
Lepmu1_308	LAESU_760772
Lindgomyces ingoldianus	LAESU_97217
Linin1_380217	Lentinula edodes
Lobaria pulmonaria	LENED_000675
Lobpul1_1077425	LENED_009785
Lobpul1_1081061	LENED_011156
Lobpul1_1086700	Leucoagaricus sp.
Lobpul1_1088690	AN958_00679
Lobpul1_1160659	AN958_01976
Lobpul1_1160823	AN958_05697
Lobpul1_1187714	AN958_05837
Lobpul1_1189558	AN958_08196
Lobpul1_1267101	AN958_09576
Lobpul1_1326505	AN958_09577
Lophiotrema nucula	AN958_11218
Lopnu1_203111	AN958_11219
Lopnu1_576877	AN958_12529
Lopnu1_603805	Moniliophthora perniciosa
Lophium mytilinum	MPER_03050
Lopmy1_551480	Moniliophthora roreri
Loramyces juncicola	Moror_10387
Lorjul_472231	Moror_10832
Lorjul_513685	Moror_11443
Loramyces macrosporus	Moror_14186
Lorma1_320020	Moror_15644
Lorma1_437337	Moror_4213
Lorma1_614065	WG66_11919
Macrophomina phaseolina	WG66_12445
Macph1_8897	WG66_17918
Macroventuria anomochaeta	WG66_18074
Macan1_446477	WG66_18690
Magnaporthe grisea	WG66_18985
Maggr1_110458	WG66_354
Maggr1_111240	WG66_8033
Mariannaea sp.	Mycena chlorophos
MarPMI226_411544	MCHLO_03985
Marssonina brunnea	MCHLO_05513

Marbr1_4753	MCHLO_07787
Massariosphaeria phaeospora	MCHLO_08688
Masph1_606827	MCHLO_13355
Melanconium sp.	Neolentinus lepideus
Melsp1_127340	NEOLE_1114180
Melsp1_95914	NEOLE_1127484
Melanomma pulvis-pyrius	NEOLE_1129527
Melpu1_277550	NEOLE_1133313
Melpu1_347683	NEOLE_1153406
Melanospora tiffanyae	NEOLE_1157631
Melti1_461564	NEOLE_1157743
Meliniomyces bicolor	NEOLE_1180214
Melbi2_645837	NEOLE_1181640
Meliniomyces variabilis	NEOLE_134104
Melva1_455976	NEOLE_318499
Metarhizium robertsii	NEOLE_467896
Metro1_2405	NEOLE_632413
Metro1_3595	Omphalotus olearius
Metro1_6916	Omp1
Metro1_9225	Omp10
Microdochium bolleyi	Omp2
Micbo1_128564	Omp3
Micbo1_13978	Omp4
Micbo1_151202	Omp5a
Micbo1_158522	Omp5b
Micbo1_181072	Omp6
Micbo1_186092	Omp7
Microdochium trichocladiopsis	Omp8
Mictri1_125659	Omp9
Mictri1_260337	Ophiostoma piceae
Mictri1_335184	F503_01342
Mictri1_375638	Ophpic1_6625
Mictri1_422579	Paxillus involutus
Microsporum canis	Paxin1_101514
Micca1_2230	Paxin1_12806
Myrothecium inundatum	Paxin1_137577
Myrin1_398933	Paxin1_167348
Myrin1_546039	Paxin1_176239
Nectria haematococca	Paxin1_180528
Necha2_74943	Paxin1_181593
Neofusicoccum parvum	Paxin1_18633
Neopa1_3315	Paxin1_77896
Neopa1_4144	Paxin1_83937
Neopa1_7973	Paxin1_86018

Neosartorya fischeri	Paxillus rubicundulus
Neofil_2116	PAXRU_23853
Niesslia exilis	PAXRU_642577
Nieex1_76034	Peniophora sp.
Oidiodendron maius	PENSP_572785
OIDMA_107833	PENSP_601208
Oidma1_107833	PENSP_625629
Ophiobolus disseminans	PENSP_626963
Ophdi1_289928	PENSP_636110
Ophdi1_418300	PENSP_682634
Ophdi1_58500	PENSP_706592
Ophiostoma novo-ulmi	PENSP_749173
Ophnu1_1985851	PENSP_755041
Paracoccidioides brasiliensis	Phanerochaete chrysosporium
Parbr1_1519	Phaca1_125341
Parbra1_1841	Phaca1_139052
Paraconiothyrium sporulosum	Phaca1_197990
Parsp1_1201140	Phaca1_211240
Parsp1_1217178	Phaca1_211244
Penicillium bilaiae	Phaca1_211256
Penbi1_460541	Phaca1_211257
Penicillium brevicompactum	Phaca1_251936
Penbr2_53488	Phaca1_259972
Penicillium canescens	Phaca1_89483
Penca1_224374	Phlebia brevispora
Penicillium chrysogenum	Phchr1_1815
Pench1_25529	Phchr1_3165
Pench1_6764	Phchr1_3229
PenchWisc1_1_144631	Phchr1_4239
Penicillium digitatum	Phchr1_4445
Pendi1_59	Phlbr1_146388
Pendi1_8028	Phlbr1_146389
Penicillium expansum	Phlbr1_148542
Penex1_331919	Phlbr1_152186
Penex1_423287	Phlbr1_153007
Penicillium janthinellum	Phlbr1_18034
Penja1_454093	Phlbr1_27358
Penicillium lanosocoeruleum	Phlbr1_71918
Penla1_395992	Phlbr1_75447
Penicillium oxalicum	Phlbr1_83077
Penox1_1709	Phlbr1_89160
Penicillium roqueforti	Phlebia centrifuga
PrARS	PHLCEN_10709
Penicillium thymicola	PHLCEN_10849

Penth1_227129	PHLCEN_10850
Periconia macrospinos	Phlebiopsis gigantea
Perma1_640487	Phlgi1_103744
Perma1_643878	Phlgi1_114823
Perma1_662832	Phlgi1_12454
Perma1_709192	Phlgi1_126738
Pestalotiopsis fici	Phlgi1_157711
PFICI_04870	Phlgi1_359064
Phaeosphaeriaceae sp.	Phlgi1_367715
PhaPMI808_630607	Phlgi1_80906
PhaPMI808_701240	Piloderma croceum
PhaPMI808_718099	Pilcr_14594
Phialocephala scopiformis	Pilcr_779936
LY89_757172	Pilcr_810716
Phisc1_722991	Pilcr_81088
Phisc1_731760	Pilcr_825684
Phisc1_779859	Pilcr_828668
Phialocephala subalpina	Pilcr_98986
PAC_01018	Pisolithus microcarpus
Phoma tracheiphila	PISMI_111694
Photr1_393361	PISMI_546554
Phyllosticta capitalensis	PISMI_636097
Phycap1_294755	PISMI_642487
Phycap1_350841	PISMI_88043
Phyllosticta citriasiana	Pisolithus tinctorius
Phycit1_361908	M404_137874
Phyllosticta citribraziliensis	M404_170039
Phcit1_228662	M404_29719
Phcit1_230456	M404_471194
Phyllosticta citricarpa	Pleurotus ostreatus
Phycitr1_625980	PleosPC15_2_1039734
Phyllosticta sp.	PleosPC15_2_1041418
Phy27169_293752	PleosPC15_2_1046456
Phy27169_350519	PleosPC15_2_1047596
Phycpc1_413892	PleosPC15_2_1048495
Phycpc1_489935	PleosPC15_2_1060726
Plectania melastoma	PleosPC15_2_1061909
Plemel1_334852	PleosPC15_2_1073415
Plemel1_353228	PleosPC15_2_1098067
Plemel1_396069	PleosPC15_2_1106708_
Plemel1_527840	PleosPC15_2_155013
Plemel1_530055	PleosPC15_2_160242
Plemel1_533333	PleosPC15_2_161354
Plectosphaerella cucumerina	PleosPC15_2_30147

Plecu1_445621	PleosPC15_2_50572
Pleomassaria siparia	Plicaturopsis crispa
Plesi1_495074	PLICR_119075
Podospora anserina	Polyporus brumalis
Podan2_5388	TPS
Podan2_5672	Postia placenta
Podospora curvicolla	POSPL_38764
Podcur1_279887	Pospl1_101754
Podcur1_310174	Pospl1_105496
Podcur1_326203	Pospl1_106438
Podcur1_408089	Pospl1_106440
Pseudographis elatina	Pospl1_125960
Pseel1_2508	Pospl1_125961
Pseudomassariella vexata	Pospl1_128412
Pseve2_338773	Pospl1_130417
Pseve2_344074	Pospl1_24705
Pseve2_354204	Pospl1_44163
Pseudovirgaria hyperparasitica	Pospl1_45581
Psehy1_445678	Pospl1_46699
Psehy1_496475	Pospl1_59374
Purpureocillium sp.	Pospl1_60326
Purspl1_260473	Pospl1_87954
Purspl1_363397	Pospl1_89105
Pyrenochaeta sp.	Pospl1_91093
Pyrsp1_595056	Pospl1_92799
Rhizoscyphus ericae	Pospl1_95481
Rhier1_616313	Pospl1_97252
Rhier1_704713	Pospl1_98072
Rhytidhysterium rufulum	Pospl1_99496
Rhyru1_1_114183	Punctularia strigosozonata
Rhyru1_1_114682	Punst1_108886
Rhyru1_1_116218	Punst1_134752
Sarcoscypha coccinea	Punst1_135766
Sarco1_413089	Punst1_136240
Sarco1_477087	Punst1_138799
Sarco1_533689	Punst1_146877
Septoria musiva	Punst1_45005
Sepmu1_150980	Punst1_61346
Sepmu1_51031	Punst1_62271
Septoria populicola	Punst1_69007
Seppo1_112324	Punst1_69869
Seppo1_36729	Pycnoporus cinnabarinus
Setosphaeria turcica	BN946_scf184637.g2
Settu1_155455	BN946_scf184747.g24

Sporothrix brasiliensis	BN946_scf184790.g3
SPBR_04258	BN946_scf184934.g16
Sporothrix schenckii	BN946_scf184940.g8
HMPREF1624_08272	BN946_scf184945.g13
Stagonospora nodorum	BN946_scf184945.g9
Stano2_10081	Rhizoctonia solani
Stano2_10963	RSOL_092870
Stagonospora sp.	RSOL_312180
Stasp1_218798	RSOL_403460
Stasp1_378012	RSOL_403680
Symbiotaphrina kochii	RSOL_510110
Symko1_913078	RSOLAG22IIIB_02130
Talaromyces marneffeii	RSOLAG22IIIB_06073
Talma1_2_9490	RSOLAG22IIIB_08057
Talaromyces proteolyticus	RSOLAG22IIIB_09566
Talpro1_398870	RSOLAG22IIIB_09570
Talaromyces stipitatus	RSOLAG22IIIB_09739
Talst1_2_11311	V565_056500
Teratosphaeria nubilosa	V565_214290
Ternu1_346415	Rhizopogon vesiculosus
Thielavia antarctica	AZE42_03256
Thian1_441220	AZE42_03257
Thielavia appendiculata	AZE42_03950
Thiap1_653559	AZE42_04671
Thielavia arenaria	AZE42_04965
Thiar1_832266	AZE42_07544
Thielavia terrestris	AZE42_08339
Thite2_2110120	AZE42_08340
Thozetella sp.	AZE42_08772
ThoPMI491_1_727832	AZE42_08877
Trematosphaeria pertusa	AZE42_10031
Trepe1_605244	AZE42_10033
Trichoderma asperellum	AZE42_12242
Trias1_142130	Rhizopogon vinicolor
Trias1_53311	K503_537004
Triasp1_109551	K503_537037
Triasp1_373402	K503_696597
Triasp1_382539	K503_699336
Trichoderma atroviride	K503_740792
Triat2_210728	K503_767681
Triat2_321366	K503_783219
Triat2_86577	K503_790659
Trichoderma citrinoviride	K503_791387
Trici4_1108149	K503_849799

Trici4_66121	Schizophyllum commune
Trichoderma guizhouense	Schco1_15679
A0028_0096870	Schco1_17515
Trichoderma harzianum	Schco1_55597
THAR02_10331	Schizopora paradoxa
Triha1_502236	SCHPA_385230
Triha1_523651	SCHPA_600612
Triha1_74633	SCHPA_600636
Trihar1_48270	SCHPA_626535
Trihar1_691238	SCHPA_825685
Trihar1_819783	SCHPA_828532
Trihar1_844963	SCHPA_828604
Trichoderma longibrachiatum	SCHPA_890331
Trilo3_1442452	SCHPA_893708
Trilo3_1456582	SCHPA_894889
Trichoderma reesei	SCHPA_910670
Trire2_112028	SCHPA_931668
Trire2_59597	SCHPA_938296
Trire2_68401	SCHPA_940716
TrireRUTC30_1_12695	SCHPA_940718
TrireRUTC30_1_75235	SCHPA_940719
Trichoderma virens	SCHPA_943858
TriviGv29_8_2_187786	SCHPA_944256
TriviGv29_8_2_222187	Scleroderma citrinum
TriviGv29_8_2_41289	SCLCI_100351
Trichophaea hybrida	SCLCI_1207283
Trihyb1_876524	SCLCI_12509
Trichophyton rubrum	SCLCI_134791
Triru1_8324	Serendipita indica
Trichophyton verrucosum	PIIN_06735
Triver1_4178	Serendipita vermifera
Trypethelium eluteriae	M408_327964
Tryvi1_496934	Serpula lacrymans
Usnea florida	cyc6_SERLA_441878
Usnflo1_55552	SerlaS7_3_2_108414
Usnflo1_574162	SerlaS7_3_2_108585
Usnflo1_877966	SerlaS7_3_2_165924
Usnflo1_901038	SerlaS7_3_2_175395
Usnflo1_955721	SerlaS7_3_2_187364
Venturia pirina	SerlaS7_3_2_61540
Venpi1_211509	SerlaS7_3_2_90456
Venpi1_218661	SerlaS7_3_2_94439
Wilcoxina mikolae	Sistotremastrum niveocreum
Wilmi1_425792	SISNI_412344

Xylaria hypoxylon	SISNI_413094
Xylhyp1_472420	SISNI_417792
Xylhyp1_503745	SISNI_419019
Xylhyp1_529710	SISNI_419037
Xylhyp1_540106	SISNI_420386
Xylhyp1_540898	SISNI_437403
Xylhyp1_549956	SISNI_445623
Xylhyp1_569642	SISNI_446492
Xylhyp1_576955	SISNI_455901
Xylhyp1_588565	SISNI_475911
Xylhyp1_614361	SISNI_482322
Xylariales sp.	SISNI_486677
XylPMI506_151792	SISNI_490653
XylPMI506_435412	SISNI_511593
XylPMI506_469434	SISNI_511679
XylPMI506_473008	SISNI_534675
XylPMI506_478051	Sistotremastrum suecicum
Zymoseptoria ardabiliae	SISSU_1009262
Zymar1_773224	SISSU_1027225
Zymoseptoria pseudotritici	SISSU_1035907
Zymps1_798041	SISSU_1035914
	SISSU_1052084
	SISSU_1061476
	SISSU_1062338
	SISSU_1062347
	SISSU_1065756
	SISSU_1067234
	SISSU_1069491
	SISSU_1132250
	SISSU_138780
	SISSU_221655
	SISSU_992550
	SISSU_993764
	Sphaerobolus stellatus
	Sphst_181402
	Sphst_184320
	Sphst_192154
	Sphst_255906
	Sphst_255948
	Sphst_266313
	Sphst_266350
	Sphst_47084
	Sphst_55620
	Sphst_68403

	Sphst_785590
	Stereum hirsutum
	STEHL_69906
	Stehi1_111121
	Stehi1_128017
	Stehi1_146390
	Stehi1_155443
	Stehi1_159379
	Stehi1_161672
	Stehi1_167646
	Stehi1_25180
	Stehi1_45387
	Stehi1_50042
	Stehi1_52743
	Stehi1_64702
	Stehi1_70268
	Stehi1_73029
	Suillus luteus
	CY34_184278
	CY34_23707
	CY34_71869
	CY34_799377
	CY34_801563
	CY34_80413
	CY34_81655
	Termitomyces sp.
	J132_01558
	J132_02641
	J132_04009
	J132_04469
	J132_04694
	J132_04698
	J132_05842
	J132_07850
	J132_08389
	J132_09198
	J132_09201
	J132_09437
	J132_09567
	J132_09570
	J132_09647
	J132_09686
	J132_09687
	J132_10181

	J132_11041
	Thanatephorus cucumeris
	BN14_00857
	BN14_03718
	RSOLAG1IB_02393
	RSOLAG1IB_05967
	RSOLAG1IB_05988
	RSOLAG1IB_06038
	Trametes pubescens
	TRAPUB_14195
	TRAPUB_4416
	TRAPUB_4417
	TRAPUB_6039
	TRAPUB_6042
	TRAPUB_7379
	TRAPUB_9141
	Trametes versicolor
	Trave1_118176
	Trave1_119121
	Trave1_122204
	Trave1_124930
	Trave1_125681
	Trave1_167198
	Trave1_169091
	Trave1_20994
	Trave1_30977
	Trave1_35003
	Trave1_44143
	Trave1_47002
	Trave1_47003
	Trave1_47026
	Trave1_75578
	Tulasnella calospora
	M407_214286
	M407_214353
	M407_49795
	M407_51027
	M407_66752
	M407_70959
	M407_78466
	Wolfiporia cocos
	Wolco1_117435
	Wolco1_120409
	Wolco1_133798

	Wolco1_134393
	Wolco1_145847
	Wolco1_150507
	Wolco1_15395
	Wolco1_162429
	Wolco1_61127
	Wolco1_62102
	Wolco1_63709
	Wolco1_70381
	Wolco1_72514
	Wolco1_72849
	Wolco1_89832
	Wolco1_95045
	Wolco1_95361

[00163] With the 1408 TPSs, a phylogenetic tree which has seven major distinct TPS clades (Figure 18) was built and all-by-all BLAST analysis with enzyme function initiative (EFI) - enzyme similarity tool (EST) (Figure 19) was carried out. All the clades have at least two characterized TPS enzymes. The promiscuous TPSs producing a series of muurolene and cadinene compounds (such as Omp1-3, Cop1-3, AAE3_13190 and AAE3_6595) clustered together in clade I (Figure 18). Interestingly, all of the characterized TPS in clade I catalysed 1,10 cyclization of FPP (Figure 17). All the eight characterized Δ^6 -protoilludene (1,11 cyclization of FPP, Figure 17) synthases, together with other 32 putative TPSs, closely grouped in clade II. In addition, the four TPSs AAE3_9008, AAE3_6743, AAE3_0444, AAE3_5024 segmented closely in clade II, with multiple products including muurolene and cadinene (1,10 cyclization of FPP, Figure 17). TPSs with cadinene as the major product (Cop4, Stehi1_128017, Omp4-5a,b and AAE3_9164) clustered together in clade III. Viridiflorene and viridiflorol synthases (AAE3_12839 and AAE3_13291, 1,11 cyclization of FPP, Figure 17) were also in clade III but were distinct from cadiene synthases. Hyp1, 2 and 5 scattered loosely in clade IV as they have different products and different catalytic mechanisms. Hyp1 produces a linear terpene nerolidol, but Hyp2 and Hyp5 catalyze cadinene and bulnesene, respectively (1,10 cyclization of FPP). Ffsc4 (koraiol, 1,11 cyclization) and some TPSs responsible for 1,6 or 1,7 cyclization of FPP (Omp8-10, Cop6 and FsTDS (trichodiene)) clustered in clade V. Moreover, Hyp4 (unknown sesquiterpene products) and the monoterpene synthase Hyp3 (1,8-cineole) segmented in clade V. In clade VI, only two aristolochene (1,10 cyclization of FPP) synthases (AtARS (Cane and Kang, 2000) and PrARS

(Hohn and Plattner, 1989)) were characterized. Lastly, a few characterized TPSs with different cyclization mechanisms, including STC3 ((+)-eremophilene, 1,10 cyclization of FPP), STC5 ((-)-guaia-6,10(14)-diene, 1,10 cyclization of FPP), BcBOT2 (presilphiperfolan-8 β -ol, 1,11 cyclization of FPP) and Ffsc6 ((-)-a-acorenol, 1,6 cyclization of FPP), scattered in clade VII. In sum, most of Basidiomycota TPSs (including all the 11 *A. aegerita* TPSs) grouped in clade I, II and III but Ascomycota TPSs mainly scattered in clade IV, V, VI and VII. Basidiomycota TPSs, especially closely clustered ones, in each clade often share the similar cyclization mechanism. In contrast, Ascomycota TPSs in the same clade could have diverse cyclization mechanisms.

[00164] **Example 6**

[00165] *Predictive Framework to Uncover Other Fungal Viridiflorol Synthases*

[00166] Knowledge acquired by studying the TPS products and the sequence conservation in each distinct clade provides a valuable basis for mechanistic understanding of the distinct activities. And it could be used to engineer and design more effective enzymes and to probe and even predict the functions of unknown TPSs. To test the predictive capability of the framework, identification of viridiflorol synthases in other species was carried out. The reason viridiflorol synthase was chosen is that there is only one type of plant viridiflorol synthase reported among all kinds of species. Analyzed by the phylogenetic tree in Figure 18 and all-by-all BLAST in Figure 19, 15 fungal TPS homologs were closely clustered. Four of them (Sphst_47084 from *Sphaerobolus stellatus*, Denbi1_816208 from *Dendrothele bispora*, Galma_104215 from *Galerina marginata* and Pilcr_825684 from *Piloderma croceum*) were recombinantly expressed in *E. coli* and their products were analysed. The TPSs highlighted with circle (“●”) were characterized in this study. Among them, four TPSs (Sphst_47084 from *Sphaerobolus stellatus*, Denbi1_816208 from *Dendrothele bispora*, Galma_104215 from *Galerina marginata* and Pilcr_825684 from *Piloderma croceum*) were cloned and expressed in the chassis strain. The *E. coli* culture expressing Sphst_47084 and Denbi1_816208, the most closely related to AAE3_13291, produced identical products as AAE3_13291, viridiflorol 7 (~90%) and viridiflorene 6 (~10%) (Figure 24). Interestingly, the main product of Galma_104215 was tentatively identified as β -gurjunene, a compound structurally similar to viridiflorene (Figure 24). However, the cells expressing Pilcr_825684

produced γ -cadinene as the main product and a few minor sesquiterpenes including viridiflorene.

The results support that the phylogenetic tree could be used for identification of novel TPSs with similar functions.

5

[00167] **Example 7**

[00168] *Prediction and validation of fungal linalool and nerolidol synthases (LNSs)*

[00169] Starting with the sequence of AAE3_9435, identification of other NLSs in different fungal species was obtained by a BLAST search in databases of the Joint Genome
10 Institute (JGI, <http://jgi.doe.gov/fungi>) and Universal Protein Resource (UniProt, <http://www.uniprot.org/>). EFI-EST analysis was carried out and a group of TPS homologues were shown to be clustered with AAE3_9435. By setting the alignment score to between 80 and 90, a smaller set of candidates were selected. With the selected cluster of TPSs in Figure 20. A focused alignment indicated that 11 fungal TPSs were clustered closely with
15 AAE3_9435 including two other TPSs from *Agrocybe aegerita*, AAE3_05024 and AAE3_04444 (SEQ ID NO: 27), three from *Agrocybe pediades* (Agrped1_689671 (SEQ ID NO: 2), Agrped1_689675 (SEQ ID NO: 3), Agrped1_820868), three from *Galerina marginata* (Galma_223690 (SEQ ID NO: 4), Galma_266794 (SEQ ID NO: 77), Galma_63556), two from *Hypholoma sublateritium* (Hypsu1_148365, Hypsu1_148385 (SEQ
20 ID NO: 5)) and M413_27416 from *Hebeloma cylindrosporum* (Figure 20). Characterized in our previous study, AAE3_05024, the most closely related TPS to AAE3_9435, seems to be a pseudogene. The main products of AAE3_04444 were γ -muurolene (33%) and β -cadinene (21%). The other TPSs in Figure 21 have not been functionally annotated. As a proof of concept, five out of the nine uncharacterized TPSs were chosen to test their functions. All
25 five TPSs give rise to at least one terpene when expressed in the *E. coli* strain overproducing IPP and DMAPP. In *E. coli* strains expressing Agrped1_689675 only linalool was found in the headspace of the culture. Similar to AAE3_9435, Agrped1_689671, Galma_223690 and Hypsu1_148385 showed a bifunctional NLS function producing nerolidol and linalool (Figure 21B). Galma_266794 clones led to the sesquiterpene germacrene D (62%,
30 validated by Cubeb essential oil) as main product and a few other sesquiterpenes including δ -cadinene (17%), γ -muurolene (8%), but no monoterpene. Due to the lack of the GPP synthase in the *E. coli* strain, the concentration of intracellular FPP was much higher than that of

intracellular GPP. Consequently, all the four TPSs (AAE3_9435, Agrped1_689671, Galma_223690 and Hypsu1_148385) produced nerolidol as the main product (88-96%) and linalool as the minor product (2-10%, Figure 21). However, when expressed in the same E. coli strain, Agrped1_689675 produced only linalool but no sesquiterpenes. The data indicated
5 that Agrped1_689675, is an exclusive monoterpene synthase and has no activity of sesquiterpene synthase thus this was named as 'Ape_LS'. Interestingly, despite with different products, all the six TPSs shared some very conserved regions in the metal-binding motif (such as 'DEYTD' and 'NDMHSYxxE' region). Figure 25 shows the conserved regions for sesquiterpene and monoterpene synthases. The 2 conserved domains are DD(E/N/Y/S)XXD
10 and NDSE. The two conserved domains served as an important pre-screening of terpene synthase homologues. Those homologues missing or having incomplete domains often have no activities and thus are excluded in our screening process.

[00170] **Example 8**

15 [00171] *Mutating the LS for a different function*

[00172] As an exclusive monoterpene synthase, it was hypothesized that a point mutation of Agrped1_689675 (SEQ ID NO: 3) could change its function and products. To test this hypothesis, a few positions where Agrped1_689675 and the rest are different were highlighted (Figure 22). The crystal structure in Figure 26 was used to guide the engineering
20 of Agrped1_689675. Among different amino acids, F204 was chosen as the first to mutate. It was found that 3 out of 6 mutants (F204D, F204G and F204R) had different product profiles. Unlike wild type produces only linalool, they produced both geranyl acetate (predicted by NIST library) and linalool (Figure 23) while the other three mutants (F204I, F204L and F204V) had no significant effects on enzyme activity and functions. More interestingly, the
25 production of geranyl acetate is inversely correlated with that of linalool (Figure 23C).

[00173] The homologue model of Agr1 (Agrped1_689675) and Agr3 (Agrped1_689671) was built based on the structure of 1,8-cineole synthase from *Streptomyces clavuligerus* (PDB ID: 5nx5, 5nx6). The binding pocket, consisting of 15 residues within 6 Å from the substrate, was determined by PyMOL software v2.1.1 and highlighted here. The models were
30 used to guide and understand the mutation of linalool/nerolidol synthases for improved selectivity or change of selectivity.

[00174] A summary of the sequence listing can be found in Table 5.

[00175] **Table 5.** Summary of sequence listing.

Name	Description	SEQ ID NO
AAE3_109435	Amino acid sequence of wild type <i>Agrocybe aegerita</i> FTPS	1
Agrped1_689671	Amino acid sequence of wild type <i>Agrocybe pediades</i> FTPS	2
Agrped1_689675	Amino acid sequence of wild type <i>Agrocybe pediades</i> FTPS	3
Galma_223690	Amino acid sequence of wild type <i>Galerina marginata</i> FTPS	4
Hypsu_148385	Amino acid sequence of wild type <i>Hypholoma sublateritium</i> FTPS	5
Ec.dxs	Amino acid sequence of wild type <i>Escherichia coli</i> DXS	6
Agrped1_689675_mut1	Amino acid sequence of genetically modified Agrped1_689675, C-terminal truncation	7
Agrped1_689675_mut2	Amino acid sequence of genetically modified Agrped1_689675, C-terminal truncation	8
Agrped1_689675_mut3	Amino acid sequence of genetically modified Agrped1_689675, C-terminal truncation	9
Agrped1_689675_mut4	Amino acid sequence of genetically modified Agrped1_689675, N-terminal truncation	10
Agrped1_689675_mut5	Amino acid sequence of genetically modified Agrped1_689675, N-terminal truncation	11
Agrped1_689675_mut6	Amino acid sequence of genetically modified Agrped1_689675, F204G	12
Agrped1_689675_mut7	Amino acid sequence of genetically modified Agrped1_689675, F204V	13
Agrped1_689675_mut8	Amino acid sequence of genetically modified Agrped1_689675, F204I	14
Agrped1_689675_mut9	Amino acid sequence of genetically modified Agrped1_689675, F204D	15
Agrped1_689675_mut10	Amino acid sequence of genetically modified Agrped1_689675, F204L	16
Agrped1_689675_mut11	Amino acid sequence of genetically modified Agrped1_689675, F204R	17
Agrped1_689675_mut12	Amino acid sequence of genetically modified Agrped1_689675, 1UP-3DW	18
Agrped1_689675_mut13	Amino acid sequence of genetically modified Agrped1_689675, 3UP-1DW	19
AAE3_109435_mut1	Amino acid sequence of genetically modified AAE3_109435, C-terminal truncation	20
Agrped1_689671_mut1	Amino acid sequence of genetically modified Agrped1_689671, C-terminal truncation	21
Galma_223690_mut1	Amino acid sequence of genetically modified	22

	Galma_223690, C-terminal truncation	
Hypsu_148385_mut1	Amino acid sequence of genetically modified Hypsu_148385, C-terminal truncation	23
Ec.dxs_SL3	Amino acid sequence of genetically modified <i>E. coli</i> DXS	24
Ec.dxs_SL5	Amino acid sequence of genetically modified <i>E. coli</i> DXS	25
AAE3_04120	Amino acid sequence of cDNA of wild type AAE3_04120 FTPS	26
AAE3_04444	Amino acid sequence of cDNA of wild type AAE3_04444 FTPS	27
AAE3_06595	Amino acid sequence of cDNA of wild type AAE3_06595 FTPS	28
AAE3_06743	Amino acid sequence of cDNA of wild type AAE3_06743 FTPS	29
AAE3_09164	Amino acid sequence of cDNA of wild type AAE3_09164 FTPS	30
AAE3_10454	Amino acid sequence of cDNA of wild type AAE3_10454 FTPS	31
AAE3_13291	Amino acid sequence of cDNA of wild type AAE3_13291 FTPS	32
AAE3_13190	Amino acid sequence of cDNA of wild type AAE3_13190 FTPS	33
AAE3_12839	Amino acid sequence of cDNA of wild type AAE3_12839 FTPS	34
AAE3_109435	Nucleic acid sequence of cDNA of wild type AAE3_109435 FTPS	35
Agrped1_689671	Nucleic acid sequence of cDNA of wild type Agrped1_689671 FTPS	36
Agrped1_689675	Nucleic acid sequence of cDNA of wild type Agrped1_689675 FTPS	37
Galma_223690	Nucleic acid sequence of cDNA of wild type Galma_223690 FTPS	38
Hypsu_148385	Nucleic acid sequence of cDNA of wild type Hypsu_148385 FTPS	39
AAE3_04120	Nucleic acid sequence of cDNA of wild type AAE3_04120 FTPS	40
AAE3_04444	Nucleic acid sequence of cDNA of wild type AAE3_04444 FTPS	41
AAE3_06595	Nucleic acid sequence of cDNA of wild type AAE3_06595 FTPS	42
AAE3_06743	Nucleic acid sequence of cDNA of wild type AAE3_06743 FTPS	43
AAE3_09164	Nucleic acid sequence of cDNA of wild type AAE3_09164 FTPS	44
AAE3_10454	Nucleic acid sequence of cDNA of wild type AAE3_10454 FTPS	45

AAE3_12839	Nucleic acid sequence of cDNA of wild type AAE3_12839 FTPS	46
AAE3_13190	Nucleic acid sequence of cDNA of wild type AAE3_13190 FTPS	47
AAE3_13291	Nucleic acid sequence of cDNA of wild type AAE3_13291 FTPS	48
AAE3_109435	Nucleic acid sequence of cDNA of wild type AAE3_109435 FTPS	49
Ec.dxs	Nucleic acid sequence of wild type <i>Escherichia coli</i> DXS	50
Ec.dxs_SL3	Nucleic acid sequence of genetically modified <i>E. coli</i> DXS	51
Ec.dxs_SL5	Nucleic acid sequence of genetically modified <i>E. coli</i> DXS	52
AAE3_109435	Nucleic acid sequence of wild type <i>Agrocybe aegerita</i> FTPS	53
Agrped1_689671	Nucleic acid sequence of wild type <i>Agrocybe pediades</i> FTPS	54
Agrped1_689675	Nucleic acid sequence of wild type <i>Agrocybe pediades</i> FTPS	55
Galma_223690	Nucleic acid sequence of wild type <i>Galerina marginata</i> FTPS	56
Hypsu_148385	Nucleic acid sequence of wild type <i>Hypholoma sublateritium</i> FTPS	57
Agrped1_689675_mut1	Nucleic acid sequence of genetically modified Agrped1_689675, C-terminal truncation	58
Agrped1_689675_mut2	Nucleic acid sequence of genetically modified Agrped1_689675, C-terminal truncation	59
Agrped1_689675_mut3	Nucleic acid sequence of genetically modified Agrped1_689675, C-terminal truncation	60
Agrped1_689675_mut4	Nucleic acid sequence of genetically modified Agrped1_689675, N-terminal truncation	61
Agrped1_689675_mut5	Nucleic acid sequence of genetically modified Agrped1_689675, N-terminal truncation	62
Agrped1_689675_mut6	Nucleic acid sequence of genetically modified Agrped1_689675, F204G	63
Agrped1_689675_mut7	Nucleic acid sequence of genetically modified Agrped1_689675, F204V	64
Agrped1_689675_mut8	Nucleic acid sequence of genetically modified Agrped1_689675, F204I	65
Agrped1_689675_mut9	Nucleic acid sequence of genetically modified Agrped1_689675, F204D	66
Agrped1_689675_mut10	Nucleic acid sequence of genetically modified Agrped1_689675, F204L	67
Agrped1_689675_mut11	Nucleic acid sequence of genetically modified Agrped1_689675, F204R	68
Agrped1_689675_mut12	Nucleic acid sequence of genetically modified	69

	Agrpel1_689675, 1UP-3DW	
Agrpel1_689675_mut13	Nucleic acid sequence of genetically modified Agrpel1_689675, 3UP-1DW	70
AAE3_109435_mut1	Nucleic acid sequence of genetically modified AAE3_109435, C-terminal truncation	71
Agrpel1_689671_mut1	Nucleic acid sequence of genetically modified Agrpel1_689671, C-terminal truncation	72
Galma_223690_mut1	Nucleic acid sequence of genetically modified Galma_223690, C-terminal truncation	73
Hypsu_148385_mut1	Nucleic acid sequence of genetically modified Hypsu_148385, C-terminal truncation	74
TPS31	Amino acid sequence of wild type <i>Solanum lycopersicum</i> FTPS	75
MqTPS1	Amino acid sequence of wild type <i>Melaleuca quinquenervia</i> FTPS	76
Galma_266794	Amino acid sequence of wild type <i>Galerina marginata</i> FTPS	77
Hyp3	Amino acid sequence of Hyp3 FTPS	78
Hyp5	Amino acid sequence of Hyp5 FTPS	79
Hyp2	Amino acid sequence of Hyp2 FTPS	80
Omp3	Amino acid sequence of Omp3 FTPS	81
Cop3	Amino acid sequence of Cop3 FTPS	82
Cop1	Amino acid sequence of Cop1 FTPS	83
Omp1	Amino acid sequence of Omp1 FTPS	84
Omp2	Amino acid sequence of Omp2 FTPS	85
Cop2	Amino acid sequence of Cop2 FTPS	86
Cop4	Amino acid sequence of Cop4 FTPS	87
Stehi_128017	Amino acid sequence of Stehi_128017 FTPS	88
Omp4	Amino acid sequence of Omp4 FTPS	89
Omp5a	Amino acid sequence of Omp5a FTPS	90
Omp5b	Amino acid sequence of Omp5b FTPS	91
AAE3_05024	Amino acid sequence of AAE3_05024 FTPS	92
AAE3_09008	Amino acid sequence of AAE3_09008 FTPS	93
AAE3_04210	Amino acid sequence of AAE3_04210 FTPS	94
Omp6	Amino acid sequence of Omp6 FTPS	95
Stehi_25180	Amino acid sequence of Stehi_25180 FTPS	96
Omp7	Amino acid sequence of Omp7 FTPS	97
Pro1	Amino acid sequence of Pro1 FTPS	98
Stehi_73029	Amino acid sequence of Stehi_73029 FTPS	99
Stehi_64702	Amino acid sequence of Stehi_64702 FTPS	100
Cop5	Amino acid sequence of Cop5 FTPS	101
Stehi_159379	Amino acid sequence of Stehi_159379 FTPS	102
Cop6	Amino acid sequence of Cop6 FTPS	103
Omp10	Amino acid sequence of Omp10 FTPS	104
Omp9	Amino acid sequence of Omp9 FTPS	105
Omp8	Amino acid sequence of Omp8 FTPS	106

Hyp3 metal binding domain 1	First metal binding domain of Hyp3 FTPS	107
Hyp3 metal binding domain 2	Second metal binding domain of Hyp3 FTPS	108
Hyp5 metal binding domain 1	First metal binding domain of Hyp5 FTPS	109
Hyp5 metal binding domain 2	Second metal binding domain of Hyp5 FTPS	110
Hyp2 metal binding domain 1	First metal binding domain of Hyp2 FTPS	111
Hyp2 metal binding domain 2	Second metal binding domain of Hyp2 FTPS	112
Omp3 metal binding domain 1	First metal binding domain of Omp3 FTPS	113
Omp3 metal binding domain 2	Second metal binding domain of Omp3 FTPS	114
AAE3_13190 metal binding domain 1	First metal binding domain of AAE3_13190 FTPS	115
AAE3_13190 metal binding domain 2	Second metal binding domain of AAE3_13190 FTPS	116
Cop3 metal binding domain 1	First metal binding domain of Cop3 FTPS	117
Cop3 metal binding domain 2	Second metal binding domain of Cop3 FTPS	118
AAE3_06595 metal binding domain 1	First metal binding domain of AAE3_06595 FTPS	119
AAE3_06595 metal binding domain 2	Second metal binding domain of AAE3_06595 FTPS	120
Cop1 metal binding domain 1	First metal binding domain of Cop1 FTPS	121
Cop1 metal binding domain 2	Second metal binding domain of Cop1 FTPS	122
Omp1 metal binding domain 1	First metal binding domain of Omp1 FTPS	123
Omp1 metal binding domain 2	Second metal binding domain of Omp1 FTPS	124
Omp2 metal binding domain 1	First metal binding domain of Omp2 FTPS	125
Omp2 metal binding domain 2	Second metal binding domain of Omp2 FTPS	126
Cop2 metal binding domain 1	First metal binding domain of Cop2 FTPS	127
Cop2 metal binding domain 2	Second metal binding domain of Cop2 FTPS	128
AAE3_12839 metal binding domain 1	First metal binding domain of AAE3_12839 FTPS	129
AAE3_12839 metal	Second metal binding domain of AAE3_12839	130

binding domain 2	FTPS	
AAE3_13291 metal binding domain 1	First metal binding domain of AAE3_13291 FTPS	131
AAE3_13291 metal binding domain 2	Second metal binding domain of AAE3_13291 FTPS	132
AAE3_09164 metal binding domain 1	First metal binding domain of AAE3_09164 FTPS	133
AAE3_09164 metal binding domain 2	Second metal binding domain of AAE3_09164 FTPS	134
Cop4 metal binding domain 1	First metal binding domain of Cop4 FTPS	135
Cop4 metal binding domain 2	Second metal binding domain of Cop4 FTPS	136
Stehi_128017 metal binding domain 1	First metal binding domain of Stehi_128017 FTPS	137
Stehi_128017 metal binding domain 2	Second metal binding domain of Stehi_128017 FTPS	138
Omp4 metal binding domain 1	First metal binding domain of Omp4 FTPS	139
Omp4 metal binding domain 2	Second metal binding domain of Omp4 FTPS	140
Omp5a metal binding domain 1	First metal binding domain of Omp5a FTPS	141
Omp5a metal binding domain 2	Second metal binding domain of Omp5a FTPS	142
Omp5b metal binding domain 1	First metal binding domain of Omp5b FTPS	143
Omp5b metal binding domain 2	Second metal binding domain of Omp5b FTPS	144
AAE3_04444 metal binding domain 1	First metal binding domain of AAE3_04444 FTPS	145
AAE3_04444 metal binding domain 2	Second metal binding domain of AAE3_04444 FTPS	146
AAE3_05024 metal binding domain 1	First metal binding domain of AAE3_05024 FTPS	147
AAE3_05024 metal binding domain 2	Second metal binding domain of AAE3_05024 FTPS	148
AAE3_06743 metal binding domain 1	First metal binding domain of AAE3_06743 FTPS	149
AAE3_06743 metal binding domain 2	Second metal binding domain of AAE3_06743 FTPS	150
AAE3_09008 metal binding domain 1	First metal binding domain of AAE3_09008 FTPS	151
AAE3_09008 metal binding domain 2	Second metal binding domain of AAE3_09008 FTPS	152
AAE3_10454 metal binding domain 1	First metal binding domain of AAE3_10454 FTPS	153

AAE3_10454 metal binding domain 2	Second metal binding domain of AAE3_10454 FTPS	154
AAE3_04210 metal binding domain 1	First metal binding domain of AAE3_04210 FTPS	155
AAE3_04210 metal binding domain 2	Second metal binding domain of AAE3_04210 FTPS	156
Omp6 metal binding domain 1	First metal binding domain of Omp6 FTPS	157
Omp6 metal binding domain 2	Second metal binding domain of Omp6 FTPS	158
Stehi_25180 metal binding domain 1	First metal binding domain of Stehi_25180 FTPS	159
Stehi_25180 metal binding domain 2	Second metal binding domain of Stehi_25180 FTPS	160
Omp7 metal binding domain 1	First metal binding domain of Omp7 FTPS	161
Omp7 metal binding domain 2	Second metal binding domain of Omp7 FTPS	162
Pro1 metal binding domain 1	First metal binding domain of Pro1 FTPS	163
Pro1 metal binding domain 2	Second metal binding domain of Pro1 FTPS	164
Stehi_73029 metal binding domain 1	First metal binding domain of Stehi_73029 FTPS	165
Stehi_73029 metal binding domain 2	Second metal binding domain of Stehi_73029 FTPS	166
Stehi_64702 metal binding domain 1	First metal binding domain of Stehi_64702 FTPS	167
Stehi_64702 metal binding domain 2	Second metal binding domain of Stehi_64702 FTPS	168
Cop5 metal binding domain 1	First metal binding domain of Cop5 FTPS	169
Cop5 metal binding domain 2	Second metal binding domain of Cop5 FTPS	170
Stehi_159379 metal binding domain 1	First metal binding domain of Stehi_159379 FTPS	171
Stehi_159379 metal binding domain 2	Second metal binding domain of Stehi_159379 FTPS	172
Cop6 metal binding domain 1	First metal binding domain of Cop6 FTPS	173
Cop6 metal binding domain 2	Second metal binding domain of Cop6 FTPS	174
Omp10 metal binding domain 1	First metal binding domain of Omp10 FTPS	175
Omp10 metal binding domain 2	Second metal binding domain of Omp10 FTPS	176
Omp9 metal binding	First metal binding domain of Omp9 FTPS	177

domain 1		
Omp9 metal binding domain 2	Second metal binding domain of Omp9 FTPS	178
Omp8 metal binding domain 1	First metal binding domain of Omp8 FTPS	179
Omp8 metal binding domain 2	Second metal binding domain of Omp8 FTPS	180

[00176] **Equivalents**

The foregoing examples are presented for the purpose of illustrating the invention and should not be construed as imposing any limitation on the scope of the invention. It will readily be
5 apparent that numerous modifications and alterations may be made to the specific embodiments of the invention described above and illustrated in the examples without departing from the principles underlying the invention. All such modifications and alterations are intended to be embraced by this application.

Claims

1. A bacterial strain comprising one or more vectors encoding
 - a) one or more enzymes to produce one or more terpene precursors; and
 - b) a fungal terpene synthase (FTPS).
- 5 2. The bacterial strain according to claim 1, wherein the one or more vectors comprise one or more nucleotide sequences encoding the one or more enzymes and the FTPS, operably linked to an inducible promoter.
3. The bacterial strain according to claims 1 and 2, wherein the one or more enzymes to produce the one or more terpene precursors is part of the 1-deoxy-D-xylulose 5-phosphate
- 10 (DXP) pathway.
4. The bacterial strain according to claim 3, wherein the enzyme is 1-deoxyxylulose-5-phosphate synthase (DXS), isopentenyl diphosphate isomerase (IDI) or both.
5. The bacterial strain according to claim 4, wherein the DXS comprises the amino acid sequence set forth in SEQ ID NO: 6.
- 15 6. The bacterial strain according to claim 4, wherein the DXS is genetically modified.
7. The bacterial strain according to claim 6, wherein the genetically modified DXS comprises the amino acid sequence set forth in SEQ ID NO: 24 or 25.
8. The bacterial strain according to claim 5 or 7, wherein the DXS comprises an amino acid sequence having at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at
- 20 least 98%, at least 99% or 100% identity to an amino acid sequence set forth in SEQ ID NO: 6, SEQ ID NO: 24 or 25.
9. The bacterial strain according to any one of claims 1 to 8, wherein the one or more enzymes to produce the one or more terpene precursors is expressed at an elevated level compared to a wild type enzyme.
- 25 10. The bacterial strain according to any one of claims 1 to 9, wherein the one or more terpene precursors is farnesyl pyrophosphate (FPP), geranyl pyrophosphate (GPP), geranylgeranyl pyrophosphate (GGPP), or combinations thereof.
11. The bacterial strain according to any one of claims 1 to 10, wherein the FTPS is a monoterpene synthase or a sesquiterpene synthase.
- 30 12. The bacterial strain according to any one of claims 1 to 11, wherein the FTPS comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO: 26, SEQ ID NO:27,

SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO: 30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 and SEQ ID NO:34.

13. The bacterial strain according to any one of claims 1 to 11, wherein the FTPS comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO:39.

14. The bacterial strain according to claim 13, wherein the FTPS is expressed at a higher level than a wild-type FTPS.

15. The bacterial strain according to any one of claims 1 to 12, wherein the FTPS is genetically modified.

16. The bacterial strain according to claim 15, wherein the genetically modified FTPS comprises an amino acid sequence selected from the group consisting of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO: 16 SEQ ID NO:17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22 and SEQ ID NO: 23 .

17. The bacterial strain according to claims 12 or 16, wherein the FTPS comprises an amino acid sequence having at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequences set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO: 16, SEQ ID NO:17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33 and SEQ ID NO: 34.

18. The bacterial strain according to any one of claims 1 to 17, wherein the bacterial strain is *Escherichia coli*.

19. A genetically modified 1-deoxyxylulose-5-phosphate synthase (DXS) enzyme, wherein the genetic modification is a mutation at one or more amino acid positions.

20. The genetically modified DXS enzyme according to claim 19, wherein the mutation is selected from the group consisting of H105T, E210D, Q459L, L415T and a combination thereof of SEQ ID NO: 6.

21. The genetically modified DXS enzyme according to claim 20, wherein the mutation is E210D, Q459L and L415T.

22. A genetically modified DXS enzyme comprising an amino acid sequence having at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequence set forth in SEQ ID NO: 24 or SEQ ID NO: 25.
23. A genetically modified fungal terpene synthase (FTPS), wherein the genetic
5 modification is a mutation at one or more amino acid positions.
24. The genetically modified FTPS according to claim 23, wherein the unmodified FTPS is isolated from *Agrocybe aegerita* or *Agrocybe pediades*.
25. The genetically modified FTPS according to claims 23 or 24, wherein the mutation is an amino acid substitution, insertion, deletion, C-terminal truncation, N-terminal truncation,
10 1UP-3DW mutation or 3UP-1DW mutation.
26. The genetically modified FTPS according to claim 25, wherein the mutation is selected from the group consisting of F204D, F204G, F204R, F204I, F204L, F204V and combinations thereof of SEQ ID NO: 3.
27. The genetically modified FTPS according to claim 23, comprising an amino acid
15 sequence having at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to an amino acid sequence selected from the group consisting of: SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22
20 and SEQ ID NO:23.
28. A method of producing a terpenoid comprising
- a) culturing the bacterial strain according to any one of claims 1 to 18 in an expression medium;
 - b) isolating the terpenoid from said expression medium.
- 25 29. A method of producing a terpenoid comprising
- a) culturing a bacterial strain comprising a vector encoding the genetically modified FTPS according to any one of claims 23 to 27 in an expression medium;
 - b) isolating the terpenoid from said expression medium.
30. The method according to claim 29, wherein the method further comprises the step of
30 isolating the FTPS from the bacterial cell and mixing the isolated FTPS with one or more terpene precursors to produce the terpenoid.

31. The method of claim 30, wherein the isolated FTPS is further purified prior to mixing with one or more terpene precursors.

32. The method of claim 31, wherein the isolated FTPS is further mixed with one or more additional enzymes prior to mixing with one or more terpene precursors.

5 33. The method of claim 32, wherein the one or more additional enzymes is Acetyl-CoA acetyltransferase, Hydroxymethylglutaryl-CoA synthase (HMGS), Hydroxymethylglutaryl-CoA synthase reductase (HMGR), IDI, mevalonate kinase (MK), Phosphomevalonate kinase (PMK) or mevalonate diphosphate decarboxylase (MVD1).

10 34. The method according to any one of claims 29 to 33, wherein the product is a monoterpenoid, sesquiterpenoid or a mixture of both.

35. A fungal terpene synthase (FTPS) encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38 and SEQ ID NO:39.

15

20

25

30