

VIRIDIFLOROL PRODUCTION IN AUXOTROPHIC *ESCHERICHIA COLI*

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of Singapore application No. 10201907354Y, filed 8 August 2019, 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 production of terpenoids and the optimization thereof.

BACKGROUND OF THE INVENTION

[0003] Terpenoids, also known as isoprenoids, have enormous regiochemical and stereochemical diversity, ranging from C5 (hemiterpenes) to C40 (tetraterpenes) to polyterpenes (such as natural rubber). 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. In accordance with their structural diversity, the functions of terpenoids range from mediating symbiotic or antagonistic interactions between organisms to electron transfer, protein prenylation, or contribution to membrane fluidity. The structural and functional diversity of terpenoids has allowed terpenoids to be widely used in various applications including pharmaceuticals, nutraceuticals, agriculture, flavorings, fragrances and biofuels.

[0004] Terpenes and terpenoids are the primary constituents of the essential oils of many plants and the main contributors to the aroma of these essential oils. One terpenoid of particular interest is viridiflorol, a known fragrance molecule isolated from Niaouli (*Melaleuca quinquenervia*) essential oil. Due to its antibacterial, anti-inflammatory and antioxidant properties, viridiflorol is currently used in several cosmetics and personal care products. In addition, viridiflorol is a strong feeding deterrent for melaleuca leaf weevil and a potent inhibitor of bacterial biofilm formation. Therefore, viridiflorol can be potentially applied in the agriculture and food industries to replace current broad-spectrum toxic pesticides and unhealthy food preservatives. However, high cost of plant extracts and supply fluctuations due to political and weather conditions have limited the production and the use of

viridiflorol. Chemical synthesis of viridiflorol has low enantioselectivity. The existence of 6 chiral centres, or 64 different stereoisomers, has posed a great challenge to the chemical synthesis of viridiflorol, resulting in low yields and ultra-high costs.

[0005] There is therefore a need to identify methods of viridiflorol synthesis that can produce higher yields.

SUMMARY

[0006] In one aspect, there is provided a method of increasing terpenoid production in a host cell that produces one or more terpenoids, comprising:

- 10 a) providing a host cell that produces one or more terpenoids, said host cell comprising a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme;
- b) modifying the vector to:
 - i. introduce an inducible promoter operably linked to the polynucleotide sequence encoding the terpene synthase enzyme; and
 - 15 ii. introduce a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site;
- c) determining the dosage of an inducer capable of inducing the inducible promoter;
- d) culturing the host cell in a culture medium in the presence of the inducer at the dosage determined from step c); and
- 20 e) isolating the terpenoid from the culture medium.

[0007] In another aspect, there is provided a host cell comprising at least one vector, wherein said at least one vector comprises:

- a) a polynucleotide sequence encoding one or more genes of the mevalonate pathway operably linked to an inducible promoter;
- 25 b) a polynucleotide sequence encoding a *ispA* gene;
- c) a polynucleotide sequence encoding a terpene synthase enzyme operably linked to an inducible promoter; and
- d) a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpene synthase enzyme compared to a wild type ribosomal binding site.
- 30

[0008] In another aspect, there is provided an *Escherichia coli* cell deficient in the genes *aroA*, *aroB* and *aroC*, comprising:

a. a first vector comprising:

i. a polynucleotide sequence encoding a viridiflorol synthase enzyme operably linked to a TM1 promoter; wherein said viridiflorol synthase enzyme is mutated, and wherein said mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24;

ii. a polynucleotide sequence encoding a RBS ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site, said polynucleotide sequence located upstream of the polynucleotide sequence encoding a viridiflorol synthase enzyme; wherein the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17;

iii. a polynucleotide sequence encoding a *ispA* gene operably linked to a promoter;

iv. a polynucleotide sequence encoding a *aroA* gene operably linked to a promoter;

b. a second vector comprising:

i. a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* operably linked to a TM2 promoter; and

ii. a polynucleotide sequence encoding a *aroC* gene operably linked to a promoter;

c. a third vector comprising:

i. a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* operably linked to a TM3 promoter; and

ii. a polynucleotide sequence encoding a *aroB* gene operably linked to a promoter.

[0009] In another aspect, there is provided an engineered viridiflorol synthase enzyme comprising one or more amino acid substitutions at positions selected from the group consisting of position 227, 249, 267, 314, 318, 326, deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24 and combinations thereof.

[0010] In another aspect, there is provided a method of producing viridiflorol comprising culturing the host cell according as described herein, or an *E.coli* cell as described herein in a

culture medium that comprises between about 0.001-0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) or between about 0.1-50 mM lactose.

[0011] In another aspect, there is provided a kit for producing viridiflorol comprising the host cell as described herein, or an *E.coli* cell as described herein, together with instructions
5 for use.

DEFINITIONS

[0012] As used herein, the term “terpene” refers to a class of organic compounds produced by plants, bacteria, fungi and insects. Terpenes are derived from isoprene, which
10 has the molecular formula C_5H_8 . Terpenes may be classified by the number of isoprene units in the molecule.

[0013] As used herein, the term “terpenoid” refers to a large and diverse class of organic compounds derived from terpenes and include terpenes. Terpenes are a subset of terpenoids. The building blocks of terpenes and terpenoids have a five-carbon isoprene unit and contain
15 additional functional groups, typically oxygen-containing functional groups. Terpenoids may be derived from the chemical modification of terpenes, such as the oxidation of terpenes or recombination of the carbon skeleton of terpenes.

[0014] As used herein, the term “terpene synthase” refers to an enzyme that enzymatically modifies the terpene precursors isopentenyl pyrophosphate (IPP), dimethylallyl
20 pyrophosphate (DMAPP) or a polyprenyl pyrophosphate, such that a terpene or terpenoid compound is produced. Terpene synthases refer to enzymes catalyzing complex carbocation-driven cyclization, rearrangement, and elimination reactions that enable the transformation of a few acyclic prenyl diphosphate substrates into a vast chemical library of terpenes or terpenoids.

[0015] The term “inducer” as used herein refers to a factor or substance that acts to stimulate the rate of transcription from an inducible promoter. Inducers may directly stimulate transcription from a promoter by physically interacting with the promoter or with
25 DNA sequences in the vicinity of the promoter. Alternatively, inducers may indirectly stimulate transcription from a promoter by interacting with molecules that, themselves, physically interact with the promoter or with DNA sequences in the vicinity of the promoter.
30 Inducers may be biomolecules or factors such as light and temperature. Examples of chemical

inducers include lactose, arabinose, tryptophan, allolactose and isopropyl β -D-1-thiogalactopyranoside (IPTG).

[0016] As used herein, the term “promoter” refers to a region of DNA that initiates transcription of a gene. A promoter may be a major promoter, a minor promoter or an alternative promoter. A major promoter is a promoter that is the most frequently used for the transcription of a gene. A promoter may be a constitutive promoter or an inducible promoter. A constitutive promoter is a promoter that is always active. The term “inducible promoter” as used herein refers to a promoter that can be regulated in the presence of inducers which may include certain biomolecules. Examples of inducible promoter systems in *Escherichia coli* include the Tet-on system, Tet-off system, T7 system, Trp system, Tac system and Lac system.

[0017] In the context of this application, the term “ribosomal binding site” refers to a site within an mRNA molecule to which a ribosome binds, allowing the ribosome to select the proper initiation codon during the initiation of translation. Ribosomal binding sites are especially critical for regulating the protein translation in prokaryotes. In some prokaryotes, this polynucleotide sequence within the mRNA is called the Shine-Dalgarno sequence, and the Shine-Dalgarno sequence base pairs with the 16S RNA of the ribosome.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] 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:

[0019] **Fig. 1** shows the biosynthetic pathway of viridiflorol/amorphadiene and its production with lactose titration. A) shows the biosynthetic pathway: module 1 AHT, including *atoB*, *hmgS* and *thmgR*; module 2 MPPI, including *mevk*, *pmk*, *pmd* and *idi*; and module 3 VI, including *vs*, *or* *ads*, and *ispA* (or *fpss*). Abbreviation for the compounds: HMG-CoA, 3-hydroxy-3-methyl-glutaryl-coenzyme A; MVA, mevalonate; MVAP, phosphomevalonate; MVAPP, diphosphomevalonate; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate. Dashed arrow indicates multiple enzymatic steps. The genes expressed encode the following enzymes: *atoB*, acetoacetyl-CoA thiolase; *hmgS*, HMG-CoA synthase; *thmgR*, truncated HMG-CoA reductase; *mevk*, mevalonate kinase; *pmk*, phosphomevalonate kinase; *pmd*, mevalonate

pyrophosphate decarboxylase; *idi*, IPP isomerase; *fpps*, FPP synthase; *vs*, viridiflorol synthase; *ads*, amorpho-4,11-diene synthase. B) shows the viridiflorol specific yields (mg/g dry cell weight (DCW)) and C) shows the titres (mg/L) over different inducer concentrations for 6 strains. D) shows the mathematical modelling of the optimal ranges of inducer dosages based on *sum***M3*. Here, '*sum*' was *M1*+*M2*+*M3*, where *M1*, *M2*, *M3* referred to the PSs of module 1, 2 and 3, respectively. E) shows mRNA data for different modules with inducer titration. Error bars, mean $\pm$ s.d., n=6.

[0020] **Fig. 2** illustrates transcriptional optimization and modelling. A) shows the specific yields (mg/g DCW) for the 27 transcriptionally tuned strains. The top 3 strains (#12, #21 and #24) were marked here. B) shows the model-predicted yields versus the experimental yields. The regression model used polynomial equation. C) shows a two-factor interaction plot. D) shows the correlation between predicted yields and *M1*, *M2* or *M3*. E) is a ternary plot of viridiflorol (mg/g DCW) versus *M1*, *M2* and *M3*.

[0021] **Fig. 3** illustrates translation optimization and correlation. A) shows the correlation between viridiflorol specific yields (mg/g DCW) and *in silico* predicted ribosomal binding site (RBS) strength or translational initiation rates (TIRs). B) shows the correlation between titre (mg/L) and TIRs. The control (#12), top 5 (#38, #39, #40, #41 and #46) and bottom 3 producer strains (#34, #35 and #31) were highlighted in the figures. C) shows RBS library I and their nucleic acid sequence. Design 1 (SEQ ID NO:60) refers to the design of RBS sequences in RBS library I. "W", the first base in Design 1, may be adenine or thymine. "D", the second base in Design 1, may be adenine, guanine or thymine. D) shows RBS library II and their nucleic acid sequence. Strain #12 and #31-52 had different upstream sequence: the upstream nucleic acid sequences of #12 and #31-52 were aacaggaggaattaaccacgacggaagcgtaaacgacactaaatagact (SEQ ID NO: 58) and actataggggaattgtgagcggataacaattcccctccactaaatagact (SEQ ID NO: 59), respectively. Design 2 (SEQ ID NO: 61) refers to the design of RBS sequences in RBS library II. "N", the first base in Design 2, may be any base. "K", the second base in Design 2, may be guanine or thymine. "R", the fifth base in Design 2, may be adenine or guanine.

[0022] **Fig. 4** shows viridiflorol synthase (VS) mutation for improved activity. A) shows the relative activities of the selected 48 single mutations based on phylogeny of fungal terpene synthases. B) shows the viridiflorol titres C) shows the specific yields of wildtype, single, double and triple mutants. Error bars, mean $\pm$ s.d., n = 3. In D), the VS homologue

model and 6 chosen amino acids mutation sites (G227C, L249F, E267S, V314Y, M318V and G326A) were highlighted in the model. The ligand pyrophosphate and magnesium are labelled. E) shows the distance (Angstrom) between the ligand pyrophosphate and the 3 amino acids (G227, E267 and V314).

5 [0023] **Fig. 5** shows the combination of truncation and mutation of VS with auxotrophic design. A) shows the primary structure of VS. The 2nd -85th amino acids and those in active sites are labelled. B) shows the viridiflorol titres in ZYM and chemically defined media. C) shows the viridiflorol specific yields in ZYM and the defined media. Error bars, mean $\pm$ s.d., n=3. Statistically significant difference of viridiflorol yield was denoted *P < 0.05 (two-tailed Student's t-test). D) shows the SDS-PAGE gel and western blot of wildtype (WT, 48 kDa) and truncated (del2-85, 39.2 kDa) VS. The control used an empty plasmid. IspA was expressed in the same operon with VS.

10 [0024] **Fig. 6** shows the fed-batch fermentation of viridiflorol and amorphadiene. A) shows the time-course profiles of dry cell weight (DCW) and viridiflorol titre (mg/L). B) shows a summary of all the methods used in the study and their viridiflorol titres achieved. C) shows time-course profiles of dry cell weight (DCW) and amorphadiene titre (mg/L). The dashed arrow refers to the induction time. Error bars, mean $\pm$ s.d., n=2.

15 [0025] **Fig. 7** shows GC chromatograms and mass spectra of viridiflorol and amorphadiene. β -Caryophyllene was spiked into the culture media as an internal control. Retention index and mass spectrum were used to validate the viridiflorol product.

20 [0026] **Fig. 8** shows the data fitting of viridiflorol yield (mg/g DCW) and lactose concentration for the 6 strains. The two equations were used for the data fitting. The relative strength of promoters for each module was listed in the figure. EQ1 was used to simulate strains #1, #7, #9, #12 and #24; EQ2 was used to simulate strain #27.

25 [0027] **Fig. 9** shows the correlation of optimal inducer dosages with different parameters (M1, M2, M3, sum and sum*M2). M1, M2, M3 referred to the PSs of module 1, 2 and 3, respectively. Sum referred to the sum of M1, M2 and M3. The optimal inducer dosages were better correlated with sum*M2 than other parameters. “■” refers to the upper boundary of optimal inducer dosages. “●” refers to the lower boundary of optimal dosages.

30 [0028] **Fig. 10** shows the validation of the inducer model. A) shows the sum*VI values for the 27 strains. B) shows the yields over different inducer dosages for the six strains (#2, #5, #13, #17, #20 and #25). C) shows the validation of the inducer model. Filled data points were

the data of the six strains for validation. “●” refers to the upper boundary of optimal inducer dosages. “■” refers to the lower boundary of optimal inducer dosages. Empty dots and squares are the first randomly chosen strains (#1, #7, #9, #12, #24 and #27 in Fig. 8). Filled dots and squares are the validating strains (#2, #5, #13, #17, #20 and #25).

- 5 [0029] **Fig. 11** shows the 4d plot and triplot of viridiflorol titre with the three modules. M1, M2, M3 referred to the promoter strengths (PS) of module 1, 2 and 3, respectively. The shade and size of the circles indicate the viridiflorol titres. Shade from dark to light: titres from low to high. Size from small to large: titres from low to high. Both figures represent the effect of the promoter strength of each module on viridiflorol titres.
- 10 [0030] **Fig. 12** shows a SDS-PAGE gel for the viridiflorol synthase (VS) with different-strength RBSs. IspA was co-expressed with VS in the same operon. VS, 48 kDa; IspA, 33 kDa.
- [0031] **Fig. 13** shows a homologue model of VS and the predicted active sites. The ligand pyrophosphate was shown in salmon colour and magnesium was in firebrick colour.
- 15 Hydrogen bonds among substrates, cofactors and active sites were shown in yellow dash lines.
- [0032] **Fig. 14** shows transmembrane topology and signal peptide prediction for VS using Phobius and Signal P 4.1 server. Both of the two webtools predicted there is a signal peptide region for VS. The S-score is signal peptide score, and those amino acids above the threshold
- 20 are predicted to be within the signal peptide region.
- [0033] **Fig. 15** shows inducer titration for viridiflorol production of the strain del2-85 (truncated VS). A) shows the titre in mg/L, B) shows the specific yield in mg/g DCW, C) shows the optical density OD600. The strain was induced by lactose or IPTG in chemically defined media.
- 25 [0034] **Fig. 16** shows the result of plasmid stability checking. The wildtype and auxotrophic strains were grown in chemically defined media with or without the use of antibiotics. After ~20 generation, the plasmid copy numbers were measured by qPCR and compared with those of the first generations. Here, the p15A-*cam-mevK-pmk-pmd-idi* (L2-5) was chosen as it was found to be the most unstable. The data here indicated the auxotrophic
- 30 strain had better stability over the wildtype strain. The qPCR primers for the plasmids were rt-mk-890-f and rt-mk-985-r. The primers for genomic DNA were *cysG*-f/r (Table 1). The

plasmid copy numbers were calculated as the ratio of the copy numbers of plasmid gene and the genomic gene.

[0035] **Fig. 17** shows plasmid maps of vectors used. A) shows a plasmid map of the vector comprising a polynucleotide sequence encoding the *atoB*, *hmgS* and truncated *hmgR* genes. B) shows a plasmid map of the vector comprising a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi*. C) shows a plasmid map of the vector comprising two genes (*ads*, amorphaadiene synthase gene and *ispA*). D) shows a plasmid map of the vector comprising a polynucleotide sequence encoding the two genes (*vs*, viridiflorol synthase gene and *ispA*).

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0036] In a first aspect, the present invention refers to a method of increasing terpenoid production in a host cell that produces one or more terpenoids, comprising:

- a) providing a host cell that produces one or more terpenoids, said host cell comprising a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme;
- b) modifying the vector to:
 - i. introduce an inducible promoter operably linked to the terpene synthase enzyme; and
 - ii. introduce a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site;
- c) determining the dosage of an inducer capable of inducing the inducible promoter;
- d) culturing the host cell in a culture medium in the presence of the inducer at the dosage determined from step c); and
- e) isolating the terpenoid from the culture medium.

[0037] In one embodiment, the vector further comprises a polynucleotide sequence encoding the *ispA* gene.

[0038] The host cell may comprise one or more additional vectors. In one embodiment, the one or more additional vectors comprise a polynucleotide sequence encoding one or more genes of the mevalonate pathway operably linked to an inducible promoter. The genes of the mevalonate pathway may include but are not limited to *atoB*, *hmgS*, truncated *hmgR*, *mevK*, *pmK*, *pmd*, and *idi*. The gene truncated *hmgR* produces a protein with a deletion of amino

acids at position 2 to 553 of *hmgR* (SEQ ID NO: 26). The genes of the mevalonate pathway may be located on one module or divided into multiple metabolic modules and each module may be expressed in a plasmid vector.

[0039] In another embodiment, the host cell further comprises two additional vectors. Each vector comprises a polynucleotide sequence encoding genes in a module. In yet another embodiment, the first additional vector comprises a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* of the mevalonate pathway and the second additional vector comprises a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* of the mevalonate pathway.

[0040] In one embodiment, the inducible promoter operably linked to the polynucleotide sequence encoding the terpene synthase enzyme may be selected from the group consisting of T7 promoter, Lac promoter, Trp promoter, Tac promoter, tetracycline (Tet) inducible promoter, LasR quorum-sensing promoter and EL222 light-inducible promoter.

[0041] In one embodiment, the inducible promoter is a T7 promoter. In another embodiment, the T7 promoter is a variant of the wild-type T7 promoter. A variant of the T7 promoter refers to a T7 promoter with a polynucleotide sequence that differs from the wild-type T7 promoter by one or more nucleotides. The T7 promoter variants are generated using mutations. The T7 promoter variant may be selected from the group consisting of TM1, TM2 and TM3. In one embodiment, TM1, TM2 and TM3 have the polynucleotide sequence of SEQ ID NO: 28, SEQ ID NO: 29 and SEQ ID NO: 30 respectively.

[0042] In one embodiment, the inducible promoter in each vector is the same T7 promoter variant. In another embodiment, the inducible promoter in each vector is a different T7 promoter variant.

[0043] In one embodiment, the inducible promoter in the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme is TM1, the inducible promoter in the first additional vector comprising the *atoB*, *hmgS* and truncated *hmgR* genes is TM2, and the inducible promoter in the second additional vector comprising the *mevK*, *pmK*, *pmd* and *idi* genes is TM3.

[0044] In one embodiment, in the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme, the polynucleotide sequence encoding the RBS is situated upstream of the polynucleotide sequence encoding the terpene synthase enzyme.

[0045] In another embodiment, the polynucleotide sequence encoding the RBS is 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: 6, 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. In yet another embodiment, the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17.

[0046] The inducer capable of inducing the inducible promoter operably linked to the terpene synthase enzyme may be selected from the group consisting of lactose, isopropyl β -D-1-thiogalactopyranoside (IPTG), tryptophan, arabinose, tetracycline, or light. In one embodiment, the inducer is lactose or IPTG.

[0047] The inducer may be added to the culture medium at the start and be present in the culture throughout the growth of the host cell. Alternatively, the inducer may be added to the culture medium after the host cell has grown to an optimal density.

[0048] In one embodiment, the inducer dosage is determined using a regression model of promoter strength and terpenoid yield. Regression is based on experimental sampling, where inducer dosages are tuned and the corresponding terpenoid yields are measured. In one embodiment, the optimum ranges for selected strains are identified and different mathematical equations are used to fit the correlation with promoter strengths of different modules.

[0049] In another embodiment, wherein the inducer dosage is determined using the product of the total promoter strength and the strength of the promoter operably linked to the *mevK*, *pmK*, *pmd* and *idi* genes. In one embodiment, where M1 refers to the promoter strength of the promoter operably linked to the *atoB*, *hmgS* and truncated *hmgR* genes, M2 refers to the promoter strength of the promoter operably linked to the *mevK*, *pmK*, *pmd* and *idi* genes, and M3 refers to the promoter strength of the promoter operably linked to the viridiflorol synthase and *ispA* gene, the inducer dosage is determined by multiplying M2 with the sum of M1, M2 and M3.

[0050] In one embodiment, the inducer dosage is between about 0.001-0.5 mM IPTG or between about 0.1-50 mM lactose. In another embodiment, the inducer dosage is between about 0.01-0.2 mM IPTG or between about 1-25 mM lactose.

[0051] In the method of increasing terpenoid production in a host cell as described herein, the host cell may be a bacterial host cell. In one embodiment, the bacterial host cell is an *Escherichia coli* cell.

[0052] The *Escherichia coli* strain may be selected from the group consisting of BL21, BL21(DE3), K-12(RV308), K-12(HMS174), K-12 (MG1655), W strain (ATCC 9637 and, JM109(DE3). In one embodiment, the *Escherichia coli* is a MG1655 strain. In another embodiment, the *Escherichia coli* is a BL21(DE3) strain.

[0053] In one embodiment, the *Escherichia coli* cell comprises a polynucleotide encoding a T7 RNA polymerase integrated into its genome. In another embodiment, the *Escherichia coli* is a MG1655 strain comprising a polynucleotide encoding a T7 RNA polymerase integrated into its genome.

[0054] In one embodiment, the host cell is deficient in at least one gene involved in amino acid synthesis.

[0055] In one embodiment, the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme and/or the one or more additional vectors further comprises a polynucleotide sequence encoding the one or more genes that the host cell is deficient in. The polynucleotide sequence encoding the one or more genes that the host cell is deficient in may be located on the same or different vector from the vector comprising the polynucleotide sequence encoding the terpene synthase enzyme. In one embodiment, the one or more genes of that the host cell is deficient in is present on only 1 vector. The one or more genes involved in amino acid synthesis that the host cell is deficient in encode one or more enzymes that produce one or more amino acids.

[0056] The one or more genes that the host cell is deficient in may be selected from the group consisting of *aroA*, *aroB* and *aroC*.

[0057] In one embodiment, the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme further comprises a polynucleotide sequence encoding the gene *aroA*. In another embodiment, the first additional vector comprising a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* further comprises a polynucleotide sequence encoding the gene *aroC*, and the second additional vector comprising a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* further comprises a polynucleotide sequence encoding the gene *aroB*.

[0058] In one embodiment, the host cell is deficient in *aroA*, *aroB* and *aroC* and the host cell comprises:

- a) a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme, a polynucleotide sequence encoding the gene *ispA* and a polynucleotide sequence encoding the gene *aroA*,
- b) a first additional vector comprising a polynucleotide sequence encoding the genes *atoB*, *hmgS*, truncated *hmgR* and *aroC*, and
- c) a second additional vector comprising a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd*, *idi* and *aroB*.

[0059] In one embodiment, the terpene synthase enzyme may be selected from the group consisting of viridiflorol synthase and amorphadiene synthase. In one embodiment, the terpene synthase enzyme is viridiflorol synthase.

[0060] In one embodiment, the viridiflorol synthase is isolated from *Agrocybe aegerita*. In another embodiment, the amorphadiene synthase is isolated from *Artemisia annua* (Sweet wormwood).

[0061] In one embodiment, the viridiflorol synthase is mutated at one or more amino acid positions. The mutation may be selected from the group consisting of substitution, insertion, deletion, truncation and combinations thereof. The mutation may be introduced by targeted mutation, random mutation, or combinations thereof. The term “truncation” when used in the context of viridiflorol synthase refers to the viridiflorol synthase sequence with the deletion of one or more amino acids.

[0062] In one embodiment, the mutation is a substitution of one or more amino acids at positions selected from the group consisting of position 227, 249, 267, 314, 318, 326 of SEQ ID NO: 24. In another embodiment, the mutation is selected from the group consisting of G227C, L249F, E267S, V314Y, M318V, G326A and combinations thereof. In another embodiment, the mutation is a deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24. In another embodiment, the mutation is selected from the group consisting of a deletion of amino acids at positions 2 to 40, 2 to 60 or 2 to 85 of SEQ ID NO: 24. In another embodiment, the mutation is selected from the group consisting of G227C, L249F, E267S, V314Y, M318V, G326A and combinations thereof and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24. In yet another embodiment, the mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24.

[0063] In one embodiment, in the method of increasing terpenoid production as described herein, the host cell is cultured in a batch, fed-batch or continuous fermentation culture. In another embodiment, the host cell is cultured in a fed-batch fermentation culture.

[0064] In another embodiment, the host cell is cultured in culture medium that does not contain antibiotics.

[0065] In one embodiment, in the method of increasing terpenoid production as described herein, the terpenoid is produced at a rate of between 0.3g/L/h to 0.5g/L/h. In another embodiment, terpenoid production is increased by at least 2000-fold compared to a host cell that produces one or more terpenoids comprising a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme that has not been modified.

[0066] In one aspect, there is provided a host cell comprising at least one vector, wherein said at least one vector comprises:

a) a polynucleotide sequence encoding one or more genes of the mevalonate pathway operably linked to an inducible promoter;

b) a polynucleotide sequence encoding a *ispA* gene;

c) a polynucleotide sequence encoding a terpene synthase enzyme operably linked to an inducible promoter; and

d) a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpene synthase enzyme compared to a wild type ribosomal binding site.

[0067] In one embodiment, the polynucleotide sequence encoding a terpene synthase enzyme and the *ispA* gene are located on the same vector, and the polynucleotide sequence encoding one or more genes of the mevalonate pathway is located on a different vector. The genes of the mevalonate pathway may include but are not limited to *atoB*, *hmgS*, truncated *hmgR*, *mevK*, *pmK*, *pmd*, and *idi*. In yet another embodiment, the polynucleotide sequence encoding a terpene synthase enzyme and the *ispA* gene are located on the same vector, and the polynucleotide sequence encoding one or more genes of the mevalonate pathway is located on one or two different vectors.

[0068] In one embodiment, the polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* of the mevalonate pathway is located on a first vector and the polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* of the mevalonate pathway is located on a second vector.

[0069] In one embodiment, the inducible promoter operably linked to polynucleotide sequence encoding one or more genes of the mevalonate pathway or to a polynucleotide sequence encoding a terpene synthase enzyme may be selected from the group consisting of T7 promoter, Lac promoter, Trp promoter, Tac promoter, tetracycline (Tet) inducible promoter, LasR quorum-sensing promoter and EL222 light-inducible promoter.

[0070] In one embodiment, the inducible promoter is a T7 promoter. In another embodiment, the T7 promoter is a variant of the wild-type T7 promoter. The T7 promoter variant may be selected from the group consisting of TM1, TM2 and TM3. In one embodiment, TM1, TM2 and TM3 have the polynucleotide sequence of SEQ ID NO: 28, SEW ID NO: 29 and SEQ ID NO: 30 respectively.

[0071] In one embodiment, the inducible promoter in each vector is the same T7 promoter variant. In another embodiment, the inducible promoter in each vector is a different T7 promoter variant.

[0072] In one embodiment, the inducible promoter in the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme is TM1, the inducible promoter in the first vector comprising the *atoB*, *hmgS* and truncated *hmgR* genes is TM2, and the inducible promoter in the second vector comprising the *mevK*, *pmK*, *pmd* and *idi* genes is TM3.

[0073] In one embodiment, in the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme, the polynucleotide sequence encoding the RBS is situated upstream of the polynucleotide sequence encoding the terpene synthase enzyme.

[0074] In another embodiment, the polynucleotide sequence encoding the RBS is 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: 6, 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. In yet another embodiment, the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17.

[0075] In one embodiment, the host cell as described herein is deficient in at least one gene involved in amino acid synthesis.

[0076] In one embodiment, the at least one vector further comprises a polynucleotide sequence encoding the one or more genes that the host cell is deficient in. The one or more genes involved in amino acid synthesis that the host cell is deficient in encode one or more enzymes that produce one or more amino acids.

5 [0077] The one or more genes that the host cell is deficient in may be selected from the group consisting of *aroA*, *aroB* and *aroC*.

[0078] In one embodiment, the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme further comprises a polynucleotide sequence encoding the gene *aroA*. In another embodiment, the first vector comprising a polynucleotide sequence
10 encoding the genes *atoB*, *hmgS* and truncated *hmgR* further comprises a polynucleotide sequence encoding the gene *aroC*, and the second vector comprising a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* further comprises a polynucleotide sequence encoding the gene *aroB*.

[0079] In one embodiment, the terpene synthase enzyme may be selected from the group
15 consisting of viridiflorol synthase and amorphadiene synthase. In one embodiment, the terpene synthase enzyme is viridiflorol synthase.

[0080] In one embodiment, the viridiflorol synthase is isolated from *Agrocybe aegerita*. In another embodiment, the amorphadiene synthase is isolated from *Artemisia annua* (Sweet wormwood).

20 [0081] In one embodiment, the viridiflorol synthase is mutated at one or more amino acid positions. The mutation may be selected from the group consisting of substitution, insertion, deletion, truncation and combinations thereof. The mutation may be introduced by targeted mutation, random mutation, or combinations thereof.

[0082] In one embodiment, the mutation is a substitution of one or more amino acids at
25 positions selected from the group consisting of position 227, 249, 267, 314, 318, 326 of SEQ ID NO: 24. In another embodiment, the mutation is selected from the group consisting of G227C, L249F, E267S, V314Y, M318V, G326A and combinations thereof. In another embodiment, the mutation is a deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24. In another embodiment, the mutation is selected from the group consisting of G227C, L249F,
30 E267S, V314Y, M318V, G326A and combinations thereof and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24. In yet another embodiment, the mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24.

[0083] The host cell as described herein may be a bacterial host cell. In one embodiment, the bacterial host cell is an *Escherichia coli* cell.

[0084] The *Escherichia coli* strain may be selected from the group consisting of BL21, BL21(DE3), K-12(RV308), K-12(HMS174), K-12 (MG1655), W strain (ATCC 9637) and JM109(DE3). In one embodiment, the *Escherichia coli* is a MG1655 strain. In another embodiment, the *Escherichia coli* is a BL21(DE3) strain.

[0085] In another aspect, there is provided an *Escherichia coli* cell deficient in the genes *aroA*, *aroB* and *aroC*, comprising:

a. a first vector comprising:

i. a polynucleotide sequence encoding a viridiflorol synthase enzyme operably linked to a TM1 promoter; wherein said viridiflorol synthase enzyme is mutated, and wherein said mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24;

ii. a polynucleotide sequence encoding a RBS ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site, said polynucleotide sequence located upstream of the polynucleotide sequence encoding a viridiflorol synthase enzyme; wherein the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17;

iii. a polynucleotide sequence encoding a *ispA* gene operably linked to a promoter;

iv. a polynucleotide sequence encoding a *aroA* gene operably linked to a promoter;

b. a second vector comprising:

i. a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* operably linked to a TM2 promoter; and

ii. a polynucleotide sequence encoding a *aroC* gene operably linked to a promoter;

c. a third vector comprising:

i. a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* operably linked to a TM3 promoter; and

ii. a polynucleotide sequence encoding a *aroB* gene operably linked to a promoter.

[0086] In another aspect, there is provided an engineered viridiflorol synthase enzyme comprising one or more amino acid substitutions at positions selected from the group consisting of position 227, 249, 267, 314, 318, 326, deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24 and combinations thereof.

5 [0087] In one embodiment, the engineered viridiflorol synthase enzyme comprises the mutation G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24.

[0088] In another aspect, there is provided a method of producing viridiflorol comprising culturing the host cell as described herein, or an *E.coli* cell as described herein in a culture medium that comprises between about 0.001-0.5 mM IPTG or between about 0.1-50 mM
10 lactose.

[0089] In one embodiment, there is provided a method of producing amorpho-4,11-diene comprising culturing the host cell as described herein in a culture medium that comprises between about 0.001-0.5 mM IPTG or between about 0.1-50 mM lactose. In one embodiment, the culture medium does not contain antibiotics.

15 [0090] In another aspect, there is provided a kit for producing viridiflorol comprising the host cell as described herein, or an *E.coli* cell as described herein, together with instructions for use.

[0091] In one embodiment, there is provided a kit for producing amorpho-4,11-diene comprising the host cell as described herein, together with instructions for use.

20

EXPERIMENTAL SECTION

[0092] 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.

25 [0093] *Materials and Methods*

[0094] Strain and plasmid construction

[0095] *E. coli* K12 MG1655 $\Delta recA \Delta endA$ DE3 and MG1655 $\Delta recA \Delta endA \Delta aroA \Delta aroB \Delta aroC$ DE3 were used for terpenoid production. The auxotrophic *E. coli* strain was obtained by sequential knockout of the three genes aroABC with the primers aroA-F&R, aroB-F&R and aroC-F&R (Table 1) and the λ red recombinase. The plasmids p15A-*spec-hmgS-atoB-hmgR* (L2-8) and p15A-*cam-mevK-pmk-pmd-idi* (L2-5) were also used. The operon *crtEBI* in the plasmid p15A-*kan-crtEBI-ispA* was replaced with the codon-

30

optimized viridiflorol synthase (VS) from *Agroclybe aegerita*, obtaining the plasmid p15A-*kan-vs-ispA*. Amorpho-4,11-diene synthase (ADS) was amplified from plasmid pACM-T7-dxs-T7-idi-ADS-*ispA* and cloned into p15A-*kan* vector together with *ispA*. The genes *aroA*, *aroB* and *aroC* were subsequently integrated into the plasmids, p15A-*kan*, p15A-*cam*, p15A-*spec*, respectively. The promoters of plasmids were mutated with in-house cloning method modified based on Agilent QuikChange II method. The strains and plasmids used in this study were summarized in Table 2.

[0096] **Table 1.** Primers used in the study

Primer name	Sequence	Remarks	SEQ ID NO
aroA-F	gtgtagagagttgagttcatggaatccctgacgttacaacccatcgctc	aroA-F	31
aroA-R	cattcaggctgcctggctaataccgcgccagctgctcgaaataatccggaa	aroA-R	32
aroC-F	cggagccgtgatggctggaaacacaattggacaactcttcgcgtaacca	aroC-F	33
aroC-R	ccagcgtggaatatcagcttcacatcggcattttgcgcccgttgccgta	aroC-R	34
aroB-F	ctgcgggtacagtaattaaggtggatgtcgcgttatggagaggattgtcg	aroB-F	35
aroB-R	ccccatttcagttcaatggcatgaccaaggtgtgtcccagattcagta	aroB-R	36
OL(24)-F	ACTAAATAGACTWDGAAGAGGCCTaaaatgcatcacatcacca	RBS library 1	37
OL(24)-R	AGGCCTCTTCHWAGTCTATTTAGTgtgctttgacgcttccgctc	RBS library 1	38
TS-20-OL(24)-F	tcactaaatagactNKAAaRgaggTAtaaaatgcatcac	RBS library 2	39
TS-20-OL(24)-R	YTTMNAGTCTATTTAGTGGAggggaattgttatccgct	RBS library 2	40
4879_DEL_F	catcaccatcaccatcacacgagccatagccccgcgcgt	VS truncation	41
4879_DEL_R	gtgatggtgatggtgatgcatttaggcctcttcttagtc	VS truncation	42

[0097] **Table 2.** Strain Nomenclature

plasmids	Remarks
p15A- <i>spec-aroC-hmgS-atoB-hmgR</i> (L2-8)	Expressing three genes in the module 1 or HAT and <i>aroC</i> gene
p15A- <i>cam-aroB-mevK-pmk-pmd-idi</i> (L2-5)	Expressing three genes in the module 2 or MPPI and <i>aroB</i> gene
p15A- <i>kan-aroA-vs-ispA</i>	Expressing three genes in the module 3 or VI (for viridiflorol) and <i>aroC</i> gene
p15A- <i>kan-aroA-ads-ispA</i>	Expressing three genes in the module 3 (for amorphadiene) and <i>aroC</i> gene
Wild-type <i>E. coli</i> MG1655 - <i>K12 MG1655 ΔrecAΔendA</i> with a T7 RNA polymerase (DE3) integrated. Auxotrophic <i>E. coli</i> MG1655Δ <i>aroABC</i> – with the three genes deleted in MG1655 wildtype. The gene VS was isolated from <i>Agroclybe aegerita</i> .	

[0098] Construction of RBS library

[0099] Two RBS libraries were designed for viridiflorol synthase (VS) using the degenerate primer OL(24)-F&R (SEQ ID NOs 37-38) and TS-20-OL(24)-F&R (SEQ ID NOs 39-40) (Table 1), respectively. The cloning method used was the Cross-Lapping In Vitro
5 Assembly (CLIVA) method. RBS design was aided by RBS Calculator, version 2.0. The predicted translation efficiencies of the RBS libraries were listed in Fig. 3C and 3D.

[00100] Statistical analysis

[00101] Linear regression models were used to study the relationships between the response (viridiflorol yields) and the variables (e.g. promoter strengths of different modules,
10 RBS strengths or TIRs of the VS gene). For transcriptional optimization, our strain design is a 3^3 full factorial design [3 factors or variables ($M1$, $M2$ and $M3$) and 3 levels ($TM1=0.92$, $TM2=0.37$, $TM3=0.16$) for each factor]. As in response surface methodology, the polynomial model (Fig. 2B) was used to describe the relationship between the response and the variables. Here, the model includes linear terms (e.g. $M1$), cross product terms (e.g. $M1*M2$) and a
15 second order term (e.g. $M1^2$) of each factor. The parameters (a , b , ..., j) were calculated by multiple linear regression. Coefficient of determination (R^2) was calculated with the polynomial model and was used to compare the fit of the models.

[00102] Enzyme engineering

[00103] Targeted point mutations were introduced into VS using in-house cloning method
20 modified based on Agilent QuikChange II method. The primers (1F&R to 52F&R) used were listed in Table 1. Random mutations were introduced with GeneMorph II Random Mutagenesis Kit (Agilent, Stratagene). Double mutation and triple mutation were introduced sequentially with the single VS mutant as the template. For VS truncation, the same cloning method was used with the primer pair 4879_DEL_F&R. All the enzyme mutants were
25 sequenced afterwards to confirm the mutation.

[00104] Homologue model building

[00105] The homologue model of VS was built on the structure of epi-isozizaene synthase (PDB ID: 4ltz). Sequence alignment (Fig. 14) was performed with Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) with default setting. The binding pocket,
30 consisting of 15 residues within 6 angstroms from the substrate, was determined by PyMOL software v2.1.1. All the protein structures in the study were prepared by PyMOL software.

[00106] Media used in the study

[00107] ZYM medium was prepared using 1 % tryptone, 0.5 % yeast extract, 25 mM Na_2HPO_4 , 25 mM KH_2PO_4 , 50 mM NH_4Cl , 5 mM Na_2SO_4 , 2 mM MgSO_4 , 0.5 % glycerol, 0.05 % glucose and 1-30 mM of α -lactose. Chemically defined medium contained 10 g/L glucose, 2 g/L $(\text{NH}_4)_2\text{SO}_4$, 4.2 g/L KH_2PO_4 , 11.24 g/L K_2HPO_4 , 1.7 g/L citric acid, 0.5 g/L MgSO_4 and 10 ml/l trace element solution, pH 7.0. The trace element solution (100X) contained 0.25 g/L $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 1.5 g/L $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, 0.15 g/L $\text{CuSO}_4 \cdot 2\text{H}_2\text{O}$, 0.3 g/L H_3BO_3 , 0.25 g/L $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.8 g/L $\text{Zn}(\text{CH}_3\text{COO})_2$, 5 g/L Fe(III) citrate and 0.84 g/L EDTA, pH 8.0.

[00108] Tube culture of the E. coli strains

[00109] For strain optimization, the cells were grown in 1 mL of ZYM medium in 14 mL BD Falcon™ tube at 28 °C/250 rpm for 2-3 days. In addition, 200 μL dodecane (with 500 mg/L β -caryophyllene as internal control) was used to extract viridiflorol during cell culture. When used the defined media, the cells were initially grown at 37 °C/250 rpm until OD₆₀₀ reached ~0.8, induced by 0.01~0.15 mM IPTG (or 1~30 mM lactose), and were then grown at 28 °C for 2 days. For wildtype strains, the media were supplemented with the antibiotics (34 $\mu\text{g}/\text{ml}$ chloramphenicol, 50 $\mu\text{g}/\text{ml}$ kanamycin and 50 $\mu\text{g}/\text{ml}$ spectinomycin) to maintain the three plasmids.

[00110] Bioreactor fermentation for viridiflorol production

[00111] Two 250 ml Mini Bioreactors (Applikon Biotechnology) were used with 100-200 mL working volume in this study. The cells (-80 °C stock) were grown 10 ml defined medium for 48h at 37 °C. Five milliliter of the grown cell suspension (OD of ~5) was harvested, washed and inoculated into 100 ml the defined medium in the bioreactor. Once OD reached about 5, feed solution (500 g/L glucose and 5 g/L MgSO_4) was added into the bioreactor at the rate of 0.6-1.8 mL/h for about 3 h and another feeding cycle was started by DO trigger or offline checking of residual glucose (e.g. DO increased abruptly or glucose concentration dropped below 3 g/L). The feeding cycle was repeated until induction. The cells were induced by 0.1 mM IPTG when OD reached about 30-40 (16-18 h from inoculation). After induction, a constant feeding rate at 5 g/L/h of glucose and 0.05 g/L/h of MgSO_4 was maintained. The culture temperature was adjusted to 30 °C and 20 % (v/v) of dodecane with 1 g/L of caryophyllene was supplemented into the bioreactor. During the fermentation, dissolved oxygen level was maintained at 30% (800 - 2000 r.p.m.) by

supplying filtered air at a gas rate of 1.5 vvm. The pH of the culture was controlled at 7.0 with 28% ammonia solution. The fed-batch experiments were performed in the defined media without any antibiotics.

[00112] Theoretic yield of viridiflorol and amorphadiene

5 [00113] It is known that the production of isoprene or isopentenyl pyrophosphate (IPP) via the mevalonate pathway under aerobic fermentation requires three acetyl coenzyme A (AcCoA), three ATP and two NAD(P)H. Therefore,

(1) $1.5 \text{ Glucose} + 2 \text{ O}_2 \rightarrow 3 \text{ AcCoA} + 3 \text{ ATP} + 3 \text{ CO}_2 + 6 \text{ NAD(P)H}$ (glycolysis)

(2) $3 \text{ AcCoA} + 2 \text{ NAD(P)H} \rightarrow \text{MVA}$

10 (3) $\text{MVA} + 3 \text{ ATP} \rightarrow \text{IPP} + \text{CO}_2$

(4) $3 \text{ IPP} \rightarrow \text{Viridiflorol/amorphadiene}$

Overall $4.5 \text{ Glucose} + 9 \text{ O}_2 \rightarrow \text{Viridiflorol/amorphadiene} + 12 \text{ CO}_2 + 15 \text{ H}_2\text{O}$, viridiflorol and amorphadiene mass yield on glucose is 27.4% and 25.2%, respectively.

[00114] Quantification of terpenoids

15 [00115] The terpenoid samples were prepared by diluting 10-20 μl of organic layer into 1000 μl hexane. The samples were analyzed on an Agilent 7890 gas chromatography equipped with an Agilent 5977B MSD. Samples were injected into Agilent VF-WAXms column with a split ratio of 40:1 at 240 °C. The oven program started at 100 °C for 1 min, was raised up to 150 °C at 50 °C/min, then to 240 °C at 15 °C/min and maintained at 240 °C
20 for another 2 min. The compound concentrations were calculated by interpolating with a standard curve prepared by authentic viridiflorol standard (Santa Cruz Biotechnology, USA) or relative ratios to the internal standard β -caryophyllene. Mass spectrometer was operated in EI mode with full scan analysis (m/z 30–300, 2 spectra/s).

[00116] RNA purification and quantitative PCR assay

25 [00117] Total RNA samples from *E. coli* cells were prepared using PureLink® RNA Mini Kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. RNA samples were collected 1 hour after induction (at an OD_{600} of ~ 1.5) in biological triplicates. RNA was reversed transcribed and cDNA was quantified in technical duplicates with SYBR green. Transcript levels were normalized by the reference *cycG*. For module 1 (HAT), the genes *hmgr* and *hmgs* were used. For module 2 (MPPI), the genes *mk* and *idi* were used; for
30 module 3 VI, the genes *vs* and *ispA* were used. All the genes were analyzed for the four

strains (#1 #7, #12, #27). The primers (primer names started with ‘rt’) used for quantitative PCR were in Table 1.

[00118] *Results*

[00119] **Example 1: Transcriptional optimization and statistical analysis**

5 [00120] The mevalonate pathway genes of *Saccharomyces cerevisiae* were grafted into the *E. coli* MG1655 strain. Briefly, the mevalonate pathway genes were divided into two metabolic modules – module 1 or AHT including the genes *atoB*, *hmgS* and truncated *hmgR* and module 2 or MPPI including the genes *mevK*, *pmK*, *pmd* and *idi* (Fig. 1). The farnesyl diphosphate (FPP) synthase (*fpps*, or *ispA*) was on module 3 or VI, together with the VS

10 gene. The three modules were expressed in three plasmids (Table 2). It has previously been shown that such module structure works for both carotenoids and sesquiterpenes. And this natural modular method by dividing genes into modules according to their positions in the metabolic pathway (e.g. upstream, midstream and downstream) enable easier diagnosis of the pathway bottlenecks. Remarkably, the fungal VS produced viridiflorol as the single product

15 (Fig. 7, where β -caryophyllene was an internal control). It has been shown that the three T7 promoter variants (TM1, TM2 and TM3, whose strengths were 92%, 37% and 16% of that of T7 promoter, respectively) are sufficient to cover an initial searching resolution. Thus, 27 different strains (strain #1 to #27, Table 3), covering a full permutation (3 promoters and 3 modules, $3^3=27$), were constructed. With different promoters, modules were coordinated

20 transcriptionally to prevent accumulation of toxic intermediates and thus maximizing the product yield.

[00121] **Table 3.** Promoter information for the 27 strains

Strain ID	Promoter			Promoter strength			Promoter indices		
	AHT	MPPI	VI	AHT	MPPI	VI	AHT	MPPI	VI
#1	TM1	TM1	TM1	92%	92%	92%	31%	31%	31%
#2	TM1	TM2	TM1	92%	37%	92%	38%	6%	38%
#3	TM1	TM3	TM1	92%	16%	92%	42%	1%	42%
#4	TM1	TM1	TM2	92%	92%	37%	38%	38%	6%
#5	TM1	TM2	TM2	92%	37%	37%	51%	8%	8%
#6	TM1	TM3	TM2	92%	16%	37%	58%	2%	9%
#7	TM1	TM1	TM3	92%	92%	16%	42%	42%	1%
#8	TM1	TM2	TM3	92%	37%	16%	58%	9%	2%
#9	TM1	TM3	TM3	92%	16%	16%	68%	2%	2%
#10	TM2	TM1	TM1	37%	92%	92%	6%	38%	38%
#11	TM2	TM2	TM1	37%	37%	92%	8%	8%	51%
#12	TM2	TM3	TM1	37%	16%	92%	9%	2%	58%

#13	TM2	TM1	TM2	37%	92%	37%	8%	51%	8%
#14	TM2	TM2	TM2	37%	37%	37%	12%	12%	12%
#15	TM2	TM3	TM2	37%	16%	37%	15%	3%	15%
#16	TM2	TM1	TM3	37%	92%	16%	9%	58%	2%
#17	TM2	TM2	TM3	37%	37%	16%	15%	15%	3%
#18	TM2	TM3	TM3	37%	16%	16%	20%	4%	4%
#19	TM3	TM1	TM1	16%	92%	92%	1%	42%	42%
#20	TM3	TM2	TM1	16%	37%	92%	2%	9%	58%
#21	TM3	TM3	TM1	16%	16%	92%	2%	2%	68%
#22	TM3	TM1	TM2	16%	92%	37%	2%	58%	9%
#23	TM3	TM2	TM2	16%	37%	37%	3%	15%	15%
#24	TM3	TM3	TM2	16%	16%	37%	4%	4%	20%
#25	TM3	TM1	TM3	16%	92%	16%	2%	68%	2%
#26	TM3	TM2	TM3	16%	37%	16%	4%	20%	4%
#27	TM3	TM3	TM3	16%	16%	16%	5%	5%	5%

[00122] Inducer concentration is critical for achieving the maximal yields of proteins or metabolites. Generally, strains with stronger promoters require less amount of inducer than those with weaker promoters. However, little is known about the absolute inducer dosage for each strain and it is tedious to test a large number of strains with many different concentrations of inducers. Therefore, a study was carried out to investigate the mathematical relationship between inducer concentrations and viridiflorol yields via auto-induction in ZYM medium with lactose as the inducer, as auto-inducing media is more convenient and can avoid human errors and mistiming of induction especially when the growth rates of different strains vary. The experiment started with six randomly selected strains (#1, #7, #9, #12, #24 and #27). Firstly, the data in Fig. 1B and 1C reaffirmed inducer dosages were critical for viridiflorol production and several strains did have distinct optimal inducer dosages. Secondly, two equations were sufficient to simulate their relationship with R^2 value of 0.85-0.95 (Fig. 8). Furthermore, the optimal ranges of inducer dosage (the upper and lower limits) determined by the data in Fig. 8 were correlated with $M1$, $M2$, $M3$, sum , and the product of sum and $M1$. Here, ' $M1$ ', ' $M2$ ' and ' $M3$ ' referred to the promoter strength (PS) of module 1, 2 and 3 of each strain, respectively; ' sum ' referred to the sum of $M1$, $M2$ and $M3$ (e.g. for strain #7, $M1$, $M2$ and $M3$ were 0.92, 0.92 and 0.16, respectively; sum was 0.92+0.92+0.16=2 and $sum*M3$ was 2*0.16=0.32). It was found the optimal inducer dosages were better correlated (inversely) with the product of sum and $M1$ (Fig. 9).

[00123] With a linear regression model, the optimal inducer ranges were predicted for all the 27 strains (Fig. 1D and Fig. 10A). Based on the model, the optimal ranges shifted

gradually from 5-15 mM to 25-30 mM for the 27 strains; and one of two concentrations (15 and 25 mM) would be within the optimal range for all the 27 strains. To validate the model, another six strains (#2, #5, #13, #17, #20 and #25) were tested and the experimental data were consistent with our prediction (Fig. 10B and 10C). Furthermore, mRNA data was measured to better understand the relationship between viridiflorol production and inducer dosages. It was found that for the majority of the gene transcription levels plateaued at 10-15 mM of lactose (Fig. 1E) and others at 25-30 mM, consistent with our inducer model in Fig. 1D.

[00124] Thus, the 27 strains were tested with their own optimal inducer dosages (15 or 25 mM). Among them, the top three producers were strain #12, #21 and #24, where #12 produced the high amount of viridiflorol (specific yield 174 ± 8 mg/g dry cell weight, or DCW, titre 275 ± 12 mg/L, Fig. 2A and Fig. 11). The polynomial model was used to simulate and to predict the response (viridiflorol yields) over a range of input parameter values ($M1$, $M2$ and $M3$). The model fitted with the experimental data ($R^2 = 0.78$, Fig. 2B). Based on the model, viridiflorol production was positively correlated with $M3$ but negatively with $M2$. The predicted optimal values for $M1$, $M2$ and $M3$ were 0.4 ± 0.1 , 0.1 ± 0.1 and 0.8 ± 0.1 , respectively (Fig. 2C and D). Within the 27 strains, strain #12 (whose $M1$, $M2$ and $M3$ were, 0.37, 0.16 and 0.92, respectively) was closest to the optimal values. Indeed, #12 had the highest production, ~ 180 mg/g DCW or 283 mg/L (Fig. 2A and Fig. 11).

[00125] **Example 2: Translational optimization of viridiflorol**

[00126] Through the statistical analysis (Fig. 2), it was observed that the viridiflorol yield was positively correlated with the transcription of module 3. This led to the hypothesis that the viridiflorol yield might be still limited by insufficient enzyme activities on module 3. Module 3 contains two enzymes, VS and ispA. IspA is a native *E. coli* enzyme with high expression and good activity, thus unlikely to be the limiting step. In contrast, as a fungal terpene synthase, VS was not even detectable in sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) gel when co-expressed with ispA in the same operon (Fig. 12). Thus, the production was more likely to be limited by insufficient VS activity. To test this hypothesis, two ribosomal binding site (RBS) libraries for VS were designed that covered a broad range of translational initiation rates (TIRs) from 4000 to 18,206 a.u. (strains #31 to #52, Fig. 3, here the TIRs were not experimental but *in silico* predicted values by RBS calculator, <https://salislab.net/software/>). Strain #12 (whose TIR

was 11,608 a.u.), the top producer among the 27 strains, was used as the parental strain for translational optimization (strains #31 to #52). In line with the hypothesis, the tuning of the VS RBS had a marked effect on the viridiflorol production (Fig. 3A and 3B). Among all the strains, the strain #38 produced the highest amount of viridiflorol (262 ± 28 mg/g DCW, 511 ± 37 mg/L), which was ~50% higher than its parental strain #12. A weak but positive correlation was observed between viridiflorol yields and the predicted TIRs ($R^2=0.5$ and 0.55 with specific yield and titres, respectively). The weak correlation could have had two possible explanations. Firstly, some of the predicted TIRs were not accurate; secondly, other unknown post-transcriptional effects may have led to the difference between experiment and prediction of VS. To test the first hypothesis, three representative groups of RBSs were chosen: #34 and #35 with low viridiflorol production and low TIRs; #52 with high viridiflorol production but low TIR; #38 and #39 with high viridiflorol yield and high TIRs. VS band (48 kDa) was not observed in the SDS-PAGE gels for any of them, despite the *ispA* (the polycistronic gene of VS) band (33 kDa) being quite distinct (Fig. 12).

[00127] **Example 3: Enzyme engineering of viridiflorol synthase**

[00128] After the study of transcriptional and translational optimization, it was observed that the insufficient VS activity limited the viridiflorol production. The synthetic biology toolboxes used to this point were unable to resolve this flux limitation beyond what had been obtained. To overcome it, enzyme engineering approaches were further explored. A homologue model of the fungal VS was built based on the structure of epi-isozizaene synthase (PDB ID: 4ltz). With pyrophosphate (POP) and magnesium (Mg) as ligand, the amino acids in the binding pocket were shown in Fig. 13. A BLAST search was carried out with VS as the template in UniProt (<https://www.uniprot.org/>) and Joint Genome Institute fungal database (JGI, <https://jgi.doe.gov/>). Fourteen putative fungal terpene synthases were identified that share high similarity with the VS, or AAE3_13291 (Fig. 14). Based on the homologue model and the alignment of the 15 enzymes, a small library was designed to target 16 sites which were positioned, either in vicinity of active sites, or the conserved regions (such as magnesium binding motifs). Instead of using saturation mutation, phylogeny-based mutation with the 15 enzymes (Fig. 4A) was chosen due to the lack of a suitable high-throughput screen. However, all the chosen 48 single mutations were less active than the wild type. Specifically, L174 was critical for VS activity, all 6 mutations (L174M, L174Y, L174V, L174I, L174T and L174F) at the positions reduced viridiflorol production by

>95%. L174 is positioned in vicinity of “DEYTD” (magnesium binding motif), hence its mutation could affect the binding of cofactor and thus reduced the activity. As the targeted mutation was not successful, random mutations were introduced to VS by error-prone PCR. By screening ~400 colonies, one VS mutant was found to have higher viridiflorol production.

5 Sequencing results indicated the mutant had two mutations G227C and V314Y. Despite that V314Y single mutation reduced the viridiflorol production by over 50%, V314Y and G227C together improved the viridiflorol titre by over 38% (Fig. 4B). Unlike V314 that is located in the substrate binding pocket, G227 is distant from the active sites (Fig. 4D and 4E). On top of this double mutant, other mutations, L249F, E267S, M318V and G326A, which were the best
10 4 mutations in Fig. 4A, were further introduced. Among the four triple mutants, V314Y-G227C-E267S increased viridiflorol specific yield by 52% over wildtype, despite a diminished final cell density (~17% lower than the wildtype: Fig. 4B and 4C). However, the mutant V314Y-G227C-M318V remarkably reduced viridiflorol yield over 90%.

[00129] During the alignment with its closely-related fungal terpene synthases, it was
15 observed that the VS had about additional 80 amino acids in N terminus over others (Fig. 14). Further study by signal peptide prediction programs SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) and Phobius (<http://phobius.binf.ku.dk/>) confirmed the possibility of N-terminal signal peptide in the VS (Fig. 15). The N-terminal signal peptide region may negatively affect the expression and would direct VS to *E. coli* membrane which
20 could reduce the contact with its substrate in cytosol. The latter hypothesis could also explain why wildtype VS was invisible in protein gel as membrane proteins are known to be less accessible. Hence, it was hypothesized that the truncation of this region could be beneficial to viridiflorol production. The 2nd-85th amino acids were then removed from N terminus of the wildtype and the double mutant G227C-V314Y (Fig. 5A). As a result, the truncation
25 markedly increased the viridiflorol titres and specific yields by about 200% and 100% for the wildtype and the mutant VS, respectively (Fig. 5B and 5C). In addition, the combination of truncation and enzyme mutation had synergistic effects on viridiflorol production in ZYM rich media. Consistent with the hypothesis that the wildtype VS was invisible in SDS-PAGE gels due to the N-terminal signal peptide sequence, the truncated version del2-85 (39.2 kDa)
30 was clearly detected and was further confirmed by western blot with anti-6x his tag antibody (Fig. 5D).

[00130] **Example 4: Auxotrophic *E. coli* and fed-batch fermentation of viridiflorol**

[00131] Complex medium is often subject to batch-to-batch variation and is expensive. Therefore, in large scale industrial fermentation, chemically defined media is favoured. More importantly, it was observed that biological replicates occasionally had very different results and it was likely to stem from plasmid instability, especially for the plasmid carrying the module MPPI based on the results of restriction enzyme mapping. To solve it, an auxotrophic *E. coli* strain was created by deleting three aromatic-amino-acid synthesis genes (*aroA*, *aroB* and *aroC*) from the genome. The genes were subsequently grafted into the three plasmids, respectively. Hence, this engineered strain was forced to maintain the plasmids as there was no amino acid in the defined media. Indeed, good stability was achieved for the auxotrophic strains even without antibiotics (Fig. 17).

[00132] All the four strains (WT, G227C-V314Y, del2-85, and the combination of del2-85 with G228C-V314Y) had higher production in the defined media as compared with the rich media. However, the synergistic effect of the truncation and the targeted mutation was no longer observed. The underlying reason was unknown and worth further study. A possible reason was that the achieved yield was no longer limited by VS activity but by other factors (e.g. cofactor availability) so that the synergistic benefit could not be visualized. The strain del2-85 had a slightly higher viridiflorol yield than the strain of del2-85 with G227-V314Y (Fig. 5b and c). The titration results indicated the specific yields gradually increased as the concentration of IPTG increased (Fig. 16). When induced by 0.1 mM IPTG, the cells produced 1.4 g/L viridiflorol and 1.4 g/L of dry cells from 10 g/L glucose.

[00133] **Example 5: Fed-batch fermentation**

[00134] The performance of del2-85 strain in fed-batch fermentation was tested using the chemically defined media. The cells were induced at the mid stage (OD₆₀₀ of 35). Within 62 hours, the strain produced 25.7 g/L of viridiflorol and 41 g/L of DCW. Hence, the productivity was about 0.42 g/L/h. The total glucose used in the whole process was 151 g/L, where it consisted of two stages: biomass stage and production stage (Fig. 6A). Viridiflorol was only produced in production stage and at this stage the glucose consumed was 115 g/L, therefore the carbon yield was 0.22 g/g glucose, about 81% of its theoretical yield of the mevalonate pathway (27.4%, 4.5 mole of glucose to 1 mole of viridiflorol). It had to be highlighted that the theoretical yield, which was calculated based on zero-growth condition, cannot practically achieved in fed-batch fermentation where cells were still growing.

Collectively with transcription, translation, enzyme, strain and process optimization, the viridiflorol titre was increased by 2860 fold, from 0.009 to 25.7 g/L (Fig. 6B).

[00135] **Example 6: Expending the knowledge gained to other isoprenoids**

[00136] To test if the transcriptional model in Fig. 2 can be used to optimize other
 5 terpenoids, amorpha-4,11-diene, the precursor to the potent antimalarial drug artemisinin, was chosen. Until recently, malaria still causes ~440,000 deaths each year. Both amorpha-4,11-diene and viridiflorol are sesquiterpenes and they share the same metabolic pathway except for the terpene synthase (Fig. 1). Thus, amorpha-4,11-diene is an ideal example to test the predictive feature of the model and the performance of the optimized strain. Without
 10 having to test many combinatorial pathway designs and inducer conditions, a strain was obtained that produced 1.2 g/L of amorpha-4,11-diene at batch fermentation. Using the same fermentation media and process as to viridiflorol, this strain produced 30 g/L of amorpha-4,11-diene in 80 hours with a carbon yield of 0.19 g/g glucose (Fig. 6C). The carbon yield achieved is ~75% of its theoretic yield (25.2%, 4.5 mole of glucose to 1 mole of amorpha-
 15 4,11-diene). Though the fermentation processes have not been fully optimized, the titres, production rates and yields ('TRY') for both viridiflorol and amorpha-4,11-diene here were comparable to the excellent results of Amyris Inc's artemisinic acid (25 g/L in 140 h) and higher than most of the previously reported 'TRY' data of terpenoid production in engineered microbes (1.5 g/L sclareol in 60 h, 323 mg/L nerolidol in 48 h) and plants.

20 [00137] **Table 4.** Summary of sequence listing.

Description	SEQ ID NO
Nucleic acid sequence of RBS of Strain #12	1
Nucleic acid sequence of RBS of Strain #31	2
Nucleic acid sequence of RBS of Strain #32	3
Nucleic acid sequence of RBS of Strain #33	4
Nucleic acid sequence of RBS of Strain #34	5
Nucleic acid sequence of RBS of Strain #35	6
Nucleic acid sequence of RBS of Strain #36	7
Nucleic acid sequence of RBS of Strain #37	8
Nucleic acid sequence of RBS of Strain #38	9
Nucleic acid sequence of RBS of Strain #39	10
Nucleic acid sequence of RBS of Strain #40	11
Nucleic acid sequence of RBS of Strain #41	12
Nucleic acid sequence of RBS of Strain #42	13
Nucleic acid sequence of RBS of Strain #43	14
Nucleic acid sequence of RBS of Strain #44	15

Nucleic acid sequence of RBS of Strain #45	16
Nucleic acid sequence of RBS of Strain #46	17
Nucleic acid sequence of RBS of Strain #47	18
Nucleic acid sequence of RBS of Strain #48	19
Nucleic acid sequence of RBS of Strain #49	20
Nucleic acid sequence of RBS of Strain #50	21
Nucleic acid sequence of RBS of Strain #51	22
Nucleic acid sequence of RBS of Strain #52	23
Amino acid sequence of wildtype VS	24
Amino acid sequence of full-length <i>hmgR</i>	25
Amino acid sequence of truncated <i>hmgR</i>	26
Nucleic acid sequence of promoter T7	27
Nucleic acid sequence of promoter TM1	28
Nucleic acid sequence of promoter TM2	29
Nucleic acid sequence of promoter TM3	30
aroA-F	31
aroA-R	32
aroC-F	33
aroC-R	34
aroB-F	35
aroB-R	36
OL(24)-F	37
OL(24)-R	38
TS-20-OL(24)-F	39
TS-20-OL(24)-R	40
4879_DEL_F	41
4879_DEL_R	42
Amino acid sequence of Mq.VS	43
Amino acid sequence of AAE3_13291	44
Amino acid sequence of Galma_229201	45
Amino acid sequence of Moror_10831 (WG66_18985)	46
Amino acid sequence of Galma_245845	47
Amino acid sequence of Galma_225678	48
Amino acid sequence of Galma_63553	49
Amino acid sequence of Galma_104215	50
Amino acid sequence of Pilcr_81088	51
Amino acid sequence of Pilcr_825684	52
Amino acid sequence of Sphst_47084	53
Amino acid sequence of Denbi1_816208	54
Amino acid sequence of Moror_4213	55
Amino acid sequence of Moror_10832 (WG66_18986)	56

Amino acid sequence of AAE3_12839	57
Upstream nucleic acid sequences of strain #12	58
Upstream nucleic acid sequence of strains #31-52	59
Nucleic acid sequence of RBS of Design 1	60
Nucleic acid sequence of RBS of Design 2	61
Nucleic acid sequence of codon optimized VS sequence with N-terminal 6-histidine tag	62
Nucleic acid sequence of truncated VS (del2-85) sequence with N-terminal 6-histidine tag	63

[00138] **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 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 method of increasing terpenoid production in a host cell that produces one or more terpenoids, comprising:
 - 5 a) providing a host cell that produces one or more terpenoids, said host cell comprising a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme;
 - b) modifying the vector to:
 - 10 i. introduce an inducible promoter operably linked to the polynucleotide sequence encoding the terpene synthase enzyme; and
 - ii. introduce a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site;
 - c) determining the dosage of an inducer capable of inducing the inducible promoter;
 - 15 d) culturing the host cell in a culture medium in the presence of the inducer at the dosage determined from step c); and
 - e) isolating the terpenoid from the culture medium.
2. The method of claim 1, wherein the host cell comprises one or more additional vectors, said one or more additional vectors comprising a polynucleotide sequence encoding one or more genes of the mevalonate pathway operably linked to an inducible promoter.
3. The method of claim 2, wherein the host cell further comprises two additional vectors,
 - 25 wherein the first additional vector comprises a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* of the mevalonate pathway and the second additional vector comprises a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* of the mevalonate pathway.
4. The method of any one of claims 1-3, wherein the inducible promoter is a T7 promoter variant selected from the group consisting of TM1, TM2 and TM3, wherein TM1, TM2 and TM3 have the polynucleotide sequence of SEQ ID NO: 28, SEQ ID NO: 29 and SEQ ID NO: 30 respectively.
5. The method of claim 4, wherein the inducible promoter in the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme is TM1, the inducible promoter in the first additional vector comprising the *atoB*, *hmgS* and truncated *hmgR* genes is TM2, and the inducible promoter in the second additional vector comprising the *mevK*, *pmK*, *pmd* and *idi* genes is TM3.

6. The method of any one of claims 1-5, wherein the polynucleotide sequence encoding the RBS is situated upstream of the polynucleotide sequence encoding the terpene synthase enzyme.
- 5 7. The method of claim 6, wherein the the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17.
- 10 8. The method of any one claims 1-7, wherein the inducer is lactose or isopropyl β -D-1-thiogalactopyranoside (IPTG).
- 15 9. The method of any one of claims 1-8, wherein the inducer dosage is determined using the product of the total promoter strength and the strength of the promoter operably linked to the *mevK*, *pmK*, *pmd* and *idi* genes.
- 20 10. The method of any one of claims 1-9, wherein the inducer dosage is between about 0.001-0.5 mM IPTG or between about 0.1-50 mM lactose.
- 25 11. The method of claim 10, wherein the host cell is an *Escherichia coli* cell.
- 30 12. The method of claim 11, wherein the *Escherichia coli* is a DE3 strain encoding a T7 RNA polymerase integrated into its genome.
- 35 13. The method of any one of claims 1-12, wherein the host cell is deficient in at least one gene involved in amino acid synthesis.
- 40 14. The method of claims 1-13, wherein the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme and the one or more additional vectors further comprises a polynucleotide sequence encoding the one or more genes that the host cell is deficient in.
15. The method of claims 13 or 14, wherein the at least one gene is selected from the group consisting of *aroA*, *aroB* and *aroC*.
16. The method of any one of claims 1-15, wherein the vector comprising a polynucleotide sequence encoding a terpene synthase enzyme further comprises a polynucleotide sequence encoding the gene *aroA*.
17. The method of any one of claims 3-16, wherein the first additional vector comprising a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* further comprises a polynucleotide sequence encoding the gene *aroC*, and wherein the second additional vector comprises a polynucleotide sequence encoding the genes *mevK*,

pmK, *pmd* and *idi* further comprises a polynucleotide sequence encoding the gene *aroB*.

- 5 18. The method of any one of claims 1-17, wherein the terpene synthase enzyme is viridiflorol synthase isolated from *Agrocybe aegerita*.
19. The method of claim 18, wherein the viridiflorol synthase is mutated at one or more amino acid positions.
- 10 20. The method of claim 19, wherein the mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24.
- 15 21. The method of any one of claims 1-20, wherein terpenoid production is increased by at least 2000 fold compared to a host cell that produces one or more terpenoids comprising a vector comprising a polynucleotide sequence encoding a terpene synthase enzyme that has not been modified.
- 20 22. A host cell comprising at least one vector, wherein said at least one vector comprises:
 - a) a polynucleotide sequence encoding one or more genes of the mevalonate pathway operably linked to an inducible promoter;
 - b) a polynucleotide sequence encoding a *ispA* gene;
 - c) a polynucleotide sequence encoding a terpene synthase enzyme operably linked to an inducible promoter; and
 - 25 d) a polynucleotide sequence encoding a ribosomal binding site (RBS) that increases translation initiation rate of the terpene synthase enzyme compared to a wild type ribosomal binding site.
- 30 23. An *Escherichia coli* cell deficient in the genes *aroA*, *aroB* and *aroC*, comprising:
 - a. a first vector comprising:
 - 35 i. a polynucleotide sequence encoding a viridiflorol synthase enzyme operably linked to a TM1 promoter; wherein said viridiflorol synthase enzyme is mutated, and wherein said mutation is G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24;
 - ii. a polynucleotide sequence encoding a RBS ribosomal binding site (RBS) that increases translation initiation rate of the terpenoid compared to a wild type ribosomal binding site, said polynucleotide sequence located upstream of the polynucleotide sequence encoding a viridiflorol synthase enzyme; wherein the polynucleotide sequence encoding the RBS is selected from the group consisting of SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11 and SEQ ID NO: 17;
 - 40 iii. a polynucleotide sequence encoding a *ispA* gene operably linked to a promoter;

- iv. a polynucleotide sequence encoding a *aroA* gene operably linked to a promoter;
- b. a second vector comprising:
- 5 i. a polynucleotide sequence encoding the genes *atoB*, *hmgS* and truncated *hmgR* operably linked to a TM2 promoter; and
- ii. a polynucleotide sequence encoding a *aroC* gene operably linked to a promoter;
- c. a third vector comprising:
- 10 i. a polynucleotide sequence encoding the genes *mevK*, *pmK*, *pmd* and *idi* operably linked to a TM3 promoter; and
- ii. a polynucleotide sequence encoding a *aroB* gene operably linked to a promoter.
- 15 24. An engineered viridiflorol synthase enzyme comprising one or more amino acid substitutions at positions selected from the group consisting of position 227, 249, 267, 314, 318, 326, deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24 and combinations thereof.
- 20 25. The engineered viridiflorol synthase enzyme of claim 24, comprising the mutation G227C, V314Y and deletion of amino acids at positions 2 to 85 of SEQ ID NO: 24.
- 25 26. A method of producing viridiflorol comprising culturing the host cell according to claim 22, or an *E.coli* cell according to claim 23 in a culture medium that comprises between about 0.001-0.5 mM IPTG or between about 0.1-50 mM lactose.
- 30 27. A kit for producing viridiflorol comprising the host cell according to claim 22, or an *E.coli* cell according to claim 23, together with instructions for use.

30

35

40