(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2011 (19.05.2011)

PCT

(10) International Publication Number
WO 2011/060044 A1

(51) **International Patent Classification**:
C12P 23/00 (2006.01) *C12N 1/00* (2006.01)

(21) **International Application Number**:
PCT/US2010/056190

(22) **International Filing Date**:
10 November 2010 (10.11.2010)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
12/615,985 10 November 2009 (10.11.2009) US

(71) **Applicants** *(for all designated States except US)*: **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). **UNIVERSITY OF MASSACHUSETTS** [US/US]; 225 Franklin Street, Boston, MA 02110 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **LEONARD, Effendi** [ID/US]; 1179 Boylston Street, No. 49, Boston, MA 02215 (US). **AJIKUMAR, Parayil, K.** [IN/US]; 227 Charles Street, No. 11, Cambridge, MA 02141 (US). **PRATHER, Kristalia, Lanett, Jones** [US/US]; 92 Elm Street, Milton, MA 02186 (US). **STEPHANOPOULOS, Gregory** [US/US]; 4 Russet Lane, Winchester, MA 01890 (US). **THAYER, Kelly** [US/US]; 10 Pocasset Avenue, Worcester, MA 01606 (US).

(74) **Agent**: **VAN AMSTERDAM, John, R.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,

[Continued on next page]

(54) **Title**: METHODS FOR MICROBIAL PRODUCTION OF TERPENOIDS

(57) **Abstract**: The invention relates to recombinant expression of terpenoid synthase enzymes and geranylgeranyl diphosphate synthase (GGPPS) enzymes in cells and the production of diterpenoids.

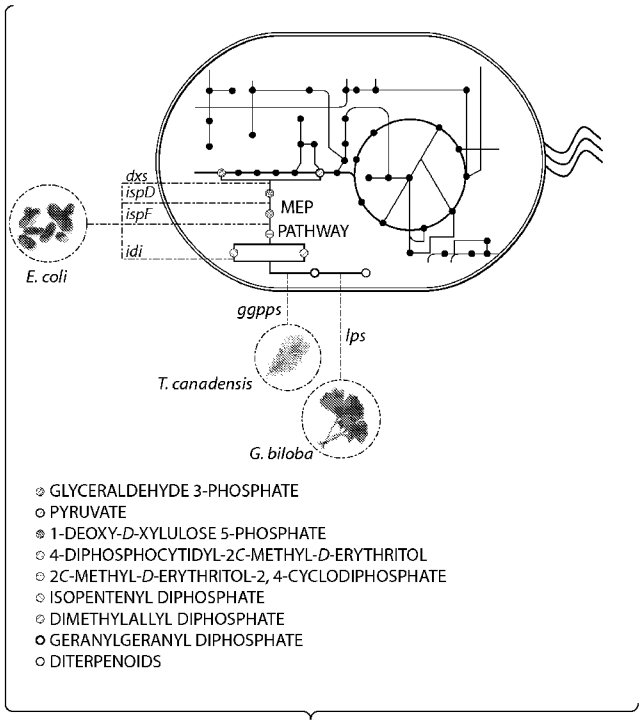


Fig. 1A

WO 2011/060044 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

METHODS FOR MICROBIAL PRODUCTION OF TERPENOIDS

Related Application

This application claims the benefit under 35 U.S.C. § 120 and 35 U.S.C. § 365(c) of U.S. Application Serial No. 12/615,985, entitled "Methods for Microbial Production of Terpenoids," filed on November 10, 2009, the entire disclosure of which is incorporated by reference herein in its entirety.

Field of the Invention

The invention relates to the production of one or more terpenoids through recombinant gene expression.

Background of the Invention

The pharmaceutically important diterpene lactone ginkgolides are products of secondary metabolism in *Ginkgo biloba* (*G. biloba*) plants. With over 3000 publications on these compounds since 2001 and annual sales of ~$250 million in the US alone, *G. biloba* extract and its constituents are currently among the most studied and sold phytochemical worldwide[1]. Ginkgolides exhibit bioactivities as antagonists of platelet-activating factor, γ-aminobutyric acid (GABA), and glycine receptors, resulting in therapeutics that are administered for improvement in vascular function, inhibition of thrombosis and embolism, and neuroprotective function[2-5]. Moreover, their potential as cancer therapeutics is under investigation[6]. Currently, the availability of ginkgolides is limited because less than 5 p.p.m of products can be obtained from leaf extract[1]. Furthermore, the growth of *G. biloba* is also extremely slow. Scalable production routes via plant cell culture and chemical synthesis have been explored; however, they are still far from industrial application. Ginkgolides yield from plant cell culture is relatively low (~40 mg/L)[7] and synthetic methods require more than 20 steps[8].

The success of fermentation technology to produce many fine and commodity chemicals has inspired the heterologous production of several plant terpenoids using microbial hosts[9-13]. In plants, secondary metabolite pathways are genetically programmed and regulated (transcriptionally and post-translationally) so that these chemicals are only synthesized as needed[14, 15]. A particular branch pathway is not designed to overproduce a certain metabolite, but rather, so that the overall metabolism works in concert. A successful

microbial production platform, on the other hand, requires that an imported pathway generate a high production yield. Metabolic engineering to increase flux through an engineered plant-derived pathway has been shown to improve terpenoid production[12, 13, 16].

Summary of the Invention

The extent of product improvement through metabolic engineering is ultimately determined by the biosynthetic capacity of the heterologous pathway in the intracellular environment of the microbial host.[17] Described herein is a novel microbial platform for producing terpenoids and diterpenoids such as levopimaradiene, the key diterpenoid precursor of the ginkgolides. This system was constructed by "tuning" a heterologous pathway to confer overproduction in a microorganism. Codon-optimized *Taxus canadensis* (*T. canadensis)* geranylgeranyl diphosphate synthase (GGPPS) and *Ginkgo biloba* (*G. biloba)* levopimaradiene synthase (LPS) were introduced into *E. coli*. To improve precursor availability, copy number of the MEP pathway in the *E. coli* host was also increased to amplify isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), the substrates of GGPPS (Fig. 1a). Diterpenoids such as levopimaradiene were successfully synthesized and their production was optimized through the generation of mutations in the LPS and GGPPS enzymes.

Aspects of the invention relate to methods that include recombinantly expressing a terpenoid synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme in a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway. In some embodiments, the cell is a bacterial cell. In certain embodiments, the cell is an *Escherichia coli* cell. In other embodiments, the cell is a Gram-positive cell such as a *Bacillus* cell. In some embodiments, the cell is a yeast cell such as a *Saccharomyces* cell or a *Yarrowia* cell. In other embodiments, the cell is an algal cell or a plant cell.

In some embodiments, the terpenoid synthase enzyme is a diterpenoid synthase enzyme such as a levopimaradiene synthase (LPS) enzyme. In some embodiments, the LPS enzyme is a *Ginkgo biloba* enzyme. In certain embodiments, the LPS enzyme contains one or more mutations. For example, the mutations in the LPS enzyme can be at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme. For example, the LPS enzyme can contain one or more mutations

selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme. In some embodiments, the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

In some embodiments, the GGPPS enzyme is a *Taxus canadensis* enzyme. In certain embodiments, the GGPPS enzyme contains one or more mutations. For example, the GGPPS enzyme can contain a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme. In certain embodiments, the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

In some embodiments, the LPS enzyme contains the mutation M593I and/or Y700F, corresponding to residues within the full-length wild-type *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme, and the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or more equivalent mutations in a homologous GGPPS enzyme.

The gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme can be expressed from one or more plasmids and/or can be incorporated into the genome of the cell. In some embodiments, the terpenoid synthase enzyme and/or the geranylgeranyl diphosphate synthase (GGPPS) enzyme is codon-optimized.

Aspects of the invention further include methods for culturing cells associated with the invention to produce a terpenoid. The terpenoids can have one or more cyclic structures. In some embodiments, the terpenoid is a diterpenoid such as levopimaradiene. Methods can further include recovering the terpenoid from the cell culture. In some embodiments, the terpenoid is recovered from the gas phase, while in other embodiments, an organic layer is

added to the cell culture, and the terpenoid is recovered from the organic layer. In some embodiments, the cell produces a Taxol, a gibberellin, and/or a steviol glycoside.

Aspects of the invention relate to cells that overexpress one or more components of the non-mevalonate (MEP) pathway, and that recombinantly express a terpenoid synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme. In some embodiments, the cell is a bacterial cell. In certain embodiments, the cell is an *Escherichia coli* cell. In other embodiments, the cell is a Gram-positive cell such as a *Bacillus* cell. In some embodiments, the cell is a yeast cell such as a *Saccharomyces* cell or a *Yarrowia* cell. In other embodiments, the cell is an algal cell or a plant cell.

In some embodiments, the terpenoid synthase enzyme is a diterpenoid synthase enzyme such as a levopimaradiene synthase (LPS) enzyme. In some embodiments, the LPS enzyme is a *Ginkgo biloba* enzyme. In certain embodiments, the LPS enzyme contains one or more mutations. For example, the mutations in the LPS enzyme can be at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme. For example, the LPS enzyme can contain one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme. In some embodiments, the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

In some embodiments, the GGPPS enzyme is a *Taxus canadensis* enzyme. In certain embodiments, the GGPPS enzyme contains one or more mutations. For example, the GGPPS enzyme can contain a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme. In certain embodiments, the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

In some embodiments, the LPS enzyme contains the mutation M593I and/or Y700F, corresponding to residues within the full-length wild-type *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme, and the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or more equivalent mutations in a homologous GGPPS enzyme.

The gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme can be expressed from one or more plasmids and/or can be incorporated into the genome of the cell. In some embodiments, the terpenoid synthase enzyme and/or the geranylgeranyl diphosphate synthase (GGPPS) enzyme is codon optimized.

In some embodiments, cells associated with the invention produce a terpenoid. The terpenoid can have one or more cyclic structures. In certain embodiments, the terpenoid is a diterpenoid such as levopimaradiene. In some embodiments, the cell produces a Taxol, a gibberellin, and/or a steviol glycoside.

Aspects of the invention relate to cells that recombinantly expresses a levopimaradiene synthase (LPS) enzyme, wherein the LPS enzyme contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme. In some embodiments, the LPS enzyme contains one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme. In certain embodiments, the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

In some embodiments, the cell is a bacterial cell. In certain embodiments, the cell is an *Escherichia coli* cell. In other embodiments, the cell is a Gram-positive cell such as a *Bacillus* cell. In some embodiments, the cell is a yeast cell such as a *Saccharomyces* cell or a

Yarrowia cell. In other embodiments, the cell is an algal cell or a plant cell. In certain embodiments, the LPS enzyme is codon optimized.

Aspects of the invention relate to cells that recombinantly expresses a geranylgeranyl diphosphate synthase (GGPPS) enzyme, wherein the GGPPS enzyme contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme. In some embodiments, the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

In some embodiments, the cell is a bacterial cell. In certain embodiments, the cell is an *Escherichia coli* cell. In other embodiments, the cell is a Gram-positive cell such as a *Bacillus* cell. In some embodiments, the cell is a yeast cell such as a *Saccharomyces* cell or a *Yarrowia* cell. In other embodiments, the cell is an algal cell or a plant cell. In certain embodiments, the LPS enzyme is codon-optimized. In certain embodiments, the GGPPS enzyme is codon-optimized.

Aspects of the invention relate to isolated levopimaradiene synthase (LPS) polypeptides that contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS polypeptide (GenBank Accession No. AF331704). For example, the isolated LPS polypeptide can contain one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V. In some embodiments, the isolated LPS polypeptide contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F. In certain embodiments, the isolated LPS polypeptide is codon-optimized. The invention also encompasses isolated nucleic acid molecule encoding any of the LPS polypeptide described herein, recombinant expression vectors comprising such nucleic acid molecules, and libraries including any of the LPS polypeptides or nucleic acid molecules described herein.

Aspects of the invention relate to isolated geranylgeranyl diphosphate synthase (GGPPS) polypeptides, wherein the GGPPS polypeptide contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis*

GGPPS polypeptide (GenBank Accession No. AF081514). In some embodiments, the isolated GGPPS polypeptide contains the mutation S239C and/or the mutation G295D. In certain embodiments, the isolated GGPPS polypeptide is codon-optimized. The invention also encompasses isolated nucleic acid molecule encoding any of the GGPPS polypeptide described herein, recombinant expression vectors comprising such nucleic acid molecules, and libraries including any of the GGPPS polypeptides or nucleic acid molecules described herein.

These and other aspects of the invention, as well as various embodiments thereof, will become more apparent in reference to the drawings and detailed description of the invention.

Brief Description of the Drawings

The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

FIG. 1 presents schematics of genetic pathways associated with aspects of the invention. FIG. 1A depicts engineering of levopimaradiene synthesis in *E. coli*. A plant-derived pathway was constructed by introducing *T. canadensis ggpps* and *G. biloba lps* codon optimized genes. To amplify the endogeneous precursor pools of GGPPS substrates (IPP and DMAPP), copy numbers of rate-limiting steps (*dxs*, *ispD*, *ispF*, *idi*) in the MEP pathway were amplified by additional episomal expression. FIG. 1B depicts a general reaction mechanism of "LPS-type" enzymes. Levopimaradiene, the major product of *G. biloba* LPS, is the gateway precursor of ginkgolides. Co-products of LPS include abietadiene, neoabiatadiene, and sandaracopimaradiene that stem from the different deprotonation patterns throughout intermediates in the reaction cascade.

FIG. 2 presents schematics and tables depicting investigation of the putative LPS binding pocket. FIG. 2A shows the homology structural model for LPS second active site. Fifteen residues within 10 Å, which were targeted for phylogenetic-based mutational analysis, are shown. The substrate in the binding pocket is farnesyl hydroxyphosphonate. FIG. 2B shows a comparison of selected residues in *G. biloba* LPS (GbLPS) with those in *A. grandis* and *P. abies* abietadiene synthase (AbAS and PaAS), and *P. abies* isopimaradiene

synthase (PaISO). Residues in GbLPS that are also found in any one of the paralogous enzymes are indicated. FIG. 2C presents a summary of LPS mutations and their impact with respect to product distribution and productivity of the engineered pathway. Trace amount (TA), not detected (ND). Numbers indicate percentage of each isomer.

FIG. 3 presents graphs depicting characteristics (productivity and product distribution) of the pre-engineered strains expressing wild-type *ggpps* and *lps* saturation mutagenesis library of: Met593 (FIG3A), Tyr700 (FIG3B), Ala620 (FIG3C), and Tyr700 (FIG3D) using *lps* M593I. In the product distribution charts, dark gray bars, light gray bars, and white bars represent proportions of levopimaradiene, abietadiene, and sandaracopimaradiene, respectively in the product mixture. Two toned (white and very light gray) bars represent nil production. WT represents the pre-engineered strain expressing wild type *ggpps* and *lps*.

FIG. 4 presents schematics and a graph depicting the generation of a GGPPS library based on stochastic mutation. FIG. 4A shows the creation of a facile high-throughput screening assay by fusing a lycopene pathway (*crt*B and *crt*I) with *ggpps* libraries. Mutant *ggpps* genes that conferred improved lycopene production (red colonies) were isolated. These variants were then co-expressed with *lps* carrying M593I/Y700F mutations for diterpenoid production assay. FIG. 4B shows the production phenotype of the pre-engineered *E. coli* strains co-expressing selected *ggpps* variants and *lps* M593I/Y700F. WT represents the strain expressing the wild-type *ggpps* and *lps* M593I/Y700F. FIG4C shows the location and identification of mutations carried by the isolated *ggpps* variants.

FIG. 5 presents graphs depicting the cultivation of an *E. coli* strain overexpressing the MEP pathway and the 'reprogrammed' plant-derived pathway constituting GGPPS S239C/G295D and LPS M593I/Y700F mutants. FIG. 5A presents diterpenoid production curves. Total diterpenoid, levopimaradiene, abietadiene, and sandaracopimaradiene are in circles, squares, triangles, and crosses, respectively. FIG. 5B depicts feed, fermentative by-product, and biomass curves. Glycerol, acetate, and cell density are in triangles, diamonds, and circles, respectively. Inverse triangle denotes the time point where 3 g of glycerol was added every 8 h.

FIG. 6 depicts GC-MS chromatograms of diterpenoid products from the engineered *E. coli* strain. FIG. 6A shows total ion chromatogram and retention time of the diterpenoids secreted in the culture media used to cultivate the pre-engineered *E. coli* strain expressing the wild type GGPPS and LPS. FIG. 6B shows GC-MS spectra of the product peaks corresponding to sandaracopimaradiene (**1**), levopimaradiene (**2**), abietadiene (**3**), and neoabietadiene (**4**) as previously reported in the literature[51,52].

FIG. 7 presents an amino acid sequence alignment of EAS[53] and LPS[54].

Detailed Description of the Invention

Aspects of the invention relate to methods and compositions for the production of one or more terpenoids through recombinant gene expression in cells. Described herein is a novel microbial platform in which a terpenoid synthase enzyme, such as levopimaradiene synthase (LPS) and a geranylgeranyl diphosphate synthase (GGPPS) enzyme are recombinantly expressed in cells. Significantly, mutations in the LPS and GGPPS enzymes have been identified herein that lead to increased production of diterpenoids. This novel microbial platform represents an unexpectedly efficient new system for producing diterpenoids such as levopimaradiene, which has widespread therapeutic applications.

This invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," or "having," "containing," "involving," and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

Aspects of the invention relate to the production of terpenoids. As used herein, a terpenoid, also referred to as an isoprenoid, is an organic chemical derived from a five-carbon isoprene unit. Several non-limiting examples of terpenoids, classified based on the number of isoprene units that they contain, include: hemiterpenoids (1 isoprene unit), monoterpenoids (2 isoprene units), sesquiterpenoids (3 isoprene units), diterpenoids (4 isoprene units), sesterterpenoids (5 isoprene units), triterpenoids (6 isoprene units), tetraterpenoids (8 isoprene units), and polyterpenoids with a larger number of isoprene units. Terpenoids are synthesized through at least two different metabolic pathways: the mevalonic acid pathway

and the MEP (2-C-methyl-D-erythritol 4-phosphate) pathway, also called the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate) pathway, the non-mevalonate pathway and the mevalonic acid-independent pathway.

Described herein are methods for producing terpenoids, such as diterpenoids, in cells through recombinant gene expression of a terpenoid synthase (also referred to as terpene cyclase) enzyme, and a geranylgeranyl diphosphate synthase (GGPPS) enzyme. In some embodiments, a terpenoid synthase enzyme is a diterpenoid synthase enzyme. Several non-limiting examples of diterpenoid synthase enzymes include casbene synthase[54], taxadiene synthase[55], levopimaradiene synthase[49], abietadiene synthase[52], isopimaradiene synthase[52], *ent*-copalyl diphosphate synthase[56], *syn*-stemar-13-ene synthase[56], *syn*-stemod-13(17)-ene synthase[56], *syn*-pimara-7,15-diene synthase[56], *ent*-sandaracopimaradiene synthase[56], *ent*-cassa-12,15-diene synthase[56], *ent*-pimara-8(14), 15-diene synthase[57], *ent*-kaur-15-ene synthase[57], *ent*-kaur-16-ene synthase[57], aphidicolan-16β-ol synthase[57], phyllocladan-16α-ol synthase[57], fusicocca-2,10(14)-diene synthase[57] and terpentetriene cyclase[58].

In some embodiments, the diterpenoid synthase enzyme is levopimaradiene synthase[49] (LPS), involved in production of levopimaradiene. In engineered systems described herein, levopimaradiene synthesis can be accompanied by production of one or more other diterpenoids such as abietadiene, sandaracopimaradiene, and neoabietadiene (trace) isomers (Fig. 1b). The GGPPS (geranylgeranyl diphosphate synthase) enzyme belongs to a prenyltransferase type family of enzymes that can accept multiple substrates (DMAPP, geranyl diphosphate (GPP), and farnesyl diphosphate (FPP)[27]). It should be appreciated that methods and compositions described herein can be used to produce a variety of different terpenoids.

According to aspects of the invention, cell(s) that recombinantly express one or more enzymes associated with the invention, and the use of such cells in producing diterpenoids such as levopimaradiene are provided. It should be appreciated that the genes encoding for the enzymes associated with the invention can be obtained from a variety of sources. In some embodiments, the gene encoding for LPS is a plant gene. For example, the gene encoding for LPS can be from a species of *Ginkgo*, such as *Ginkgo biloba* (*G. biloba*). In some embodiments, the gene encoding for GGPPS is a plant gene. For example, the gene encoding for GGPPS can be from a species of *Taxus* such as *Taxus canadensis* (*T. canadensis*). Sequences representing the wild-type DNA and protein for *G. biloba* LPS are provided by GenBank Accession No. AF331704 (SEQ ID NO:1) and AAS89668 (SEQ ID NO:2)

respectively. Sequences representing the wild-type DNA and protein for *T. canadensis* GGPPS are represented by GenBank Accession No. AF081514 (SEQ ID NO:3) and AAD16018 (SEQ ID NO:4) respectively. It should be appreciated that any of the nucleic acids and/or polypeptides described herein can be codon-optimized and expressed recombinantly in a codon-optimized form. Codon-optimized DNA and protein sequences for *T. canadensis* GGPPS are provided by SEQ ID NOs:143 and 144 respectively. Codon-optimized DNA and protein sequences for *G. biloba* LPS are provided by SEQ ID NOs:145 and 146 respectively.

As one of ordinary skill in the art would be aware, homologous genes for these enzymes can be obtained from other species and can be identified by homology searches, for example through a protein BLAST search, available at the National Center for Biotechnology Information (NCBI) internet site (www.ncbi.nlm.nih.gov). Genes associated with the invention can be cloned, for example by PCR amplification and/or restriction digestion, from DNA from any source of DNA which contains the given gene. In some embodiments, a gene associated with the invention is synthetic. Any means of obtaining a gene encoding for an enzyme associated with the invention is compatible with the instant invention.

Aspects of the invention include strategies to optimize production of a diterpenoid from a cell. Optimized production of a diterpenoid refers to producing a higher amount of a diterpenoid following pursuit of an optimization strategy than would be achieved in the absence of such a strategy. Optimization of production of a diterpenoid can involve modifying a gene encoding for an enzyme before it is recombinantly expressed in a cell. In some embodiments, such a modification involves codon optimization for expression in a bacterial cell. Codon usages for a variety of organisms can be accessed in the Codon Usage Database (www.kazusa.or.jp/codon/). Codon optimization, including identification of optimal codons for a variety of organisms, and methods for achieving codon optimization, are familiar to one of ordinary skill in the art, and can be achieved using standard methods.

In some embodiments, modifying a gene encoding for an enzyme before it is recombinantly expressed in a cell involves making one or more mutations in the gene encoding for the enzyme before it is recombinantly expressed in a cell. For example, a mutation can involve a substitution or deletion of a single nucleotide or multiple nucleotides. In some embodiments, a mutation of one or more nucleotides in a gene encoding for an enzyme will result in a mutation in the enzyme, such as a substitution or deletion of one or more amino acids.

In some embodiments "rational design" is involved in constructing specific mutations in enzymes. As used herein, "rational design" refers to incorporating knowledge of the enzyme, or related enzymes, such as its three dimensional structure, its active site(s), its substrate(s) and/or the interaction between the enzyme and substrate, into the design of the specific mutation. Based on a rational design approach, mutations can be created in an enzyme which can then be screened for increased production of a diterpenoid.

For example, as described in Example 1, rational design was implemented in creating specific mutations in LPS. Although the crystal structure of LPS is not available, the tertiary folds of other related terpene cyclase enzymes are similar. The structure of one such enzyme, 5-epi-aristolochene synthase[30] (EAS) was used to examine the second active site of LPS. This process of constructing an atomic-resolution model of one protein (e.g., LPS) from its amino acid sequence and a three-dimensional structure of a related homologous protein (e.g., EAS) is termed "homology modeling". Mutations in the second active site within other terpene cyclases impacts their plasticity[26, 31-33]. In the second active site of an LPS-type enzyme, the bicylic (+)-copalyl diphosphate (CPP) intermediate (derived from the deprotonation of GGPP in the first active site) undergoes a diphosphate-ionization cyclization. The resulting C8-sandaracopimarenyl cation intermediate is further deprotonated at two alternative sites to release isopimaradiene or sandaracopimaradiene end products. This intermediate can also undergo intramolecular proton transfer and 1,2-methyl migration to yield abietenyl cation. Subsequent deprotonation of abietenyl cation at four possible sites then produce abietadiene, levopimaradiene, neoabietadiene, and palustradiene[28, 29].

Based on the structural data, mutations in LPS were generated in fifteen residues within a 10 Å solvation layer of the LPS model: M593, C618, L619, A620, L696, Y700, K723, A727, A729, V731, N769, E777, N838, G854 and I855 (See Fig. 2a). Amino acid residue numbers indicated herein for LPS are based on amino acid numbers in the full-length, wild-type *G. biloba* LPS polypeptide (GenBank Accession No. AAS89668). One of ordinary skill in the art would understand, based on protein alignments between *G. biloba* LPS and LPS from other species, how to determine equivalent residues in other species. It should be appreciated that mutations can also be generated in other residues that are further away from the LPS active site; selection of residues for making such mutations also can be guided by homology modeling. In creating amino acid substitutions within LPS, the sequences of phylogenetically-related enzymes, such as *Abies grandis* abietadiene synthase (AS), *Picea abies* abietadiene synthase (AS) and *Picea abies* isopimardiene synthase (ISO) can be

examined and mutations in LPS can be created based, at least in part, on the sequences of these phylogenetically-related enzymes. Non-limiting examples of specific LPS mutations that can be used alone or in combination in methods associated with the invention include: M593I, M593L, M593C, M593S, M593T C618N, L619F, A620T, L696Q, Y700H, Y700F, Y700M, Y700W, K723S, A727S, A729G, V731L, N769A, E777A, N838E, G854T and I855L. It should be appreciated the methods and compositions of the invention also encompass other amino acid substitutions at these fifteen residues, as well as specific substitutions within other residues with proximity to active sites in LPS.

In some embodiments, the LPS enzyme contains a mutation in residue M593, alone or in combination with one or more other mutations. For example, the mutation can be M593I or a substitution with another hydrophobic residue such as leucine (M593L). In certain embodiments, the mutation in M593 can be M593C, M593S or M593T. Based on structural data, Met593 is located at the posterior of the binding pocket of LPS. Without wishing to be bound by any theory, hydrophobic amino acid substitutions at Met593 may improve the diterpenoid yield by disrupting hydrogen bonding at the end of the binding pocket, thus increasing the flexibility of the cavity to better fit the CPP substrate. Additionally, substitutions with large and/or bulky amino acids at Met593 may obstruct the cyclization pocket, reducing diterpenoid yield. Thus, in some embodiments, hydrophobic and/or small residues are preferred for substitution at Met593.

In some embodiments, the LPS enzyme contains a mutation in residue Y700, alone or in combination with one or more other mutations. For example, the mutation can be Y700H, Y700F, Y700M or Y700W. Based on structural data, Y700 is positioned at the entrance of the binding pocket of the enzyme, in close vicinity of a DDXXD magnesium binding motif. Without wishing to be bound by any theory, absence of a hydroxyl group in amino acids that are similar to tyrosine may allow the repositioning of the magnesium closer to the aspartate-rich region, potentially increasing reaction efficiency by improving the chelation of the diphosphate group.

In some embodiments, the LPS enzyme contains a mutation in residue A620, alone or in combination with one or more other mutations. In some embodiments, the mutation involves a substitution with a residue that is small and/or hydrophilic. In certain embodiments, the mutation can be A620C, A620G, A620S or A620T.

The LPS enzyme can contain one mutation or multiple mutations. In some embodiments, the LPS enzyme contains a mutation in M593 and a mutation in Y700. For

example, the LPS enzyme can contain the following combinations of mutations: M593I and Y700F, M593I and Y700A, or M593I and Y700C. The LPS enzyme containing these mutations can also contain one or more other mutations.

In some embodiments, random mutagenesis is used for constructing specific mutations in enzymes. As described in Example 1, improved diterpenoid production was achieved in part through random mutagenesis of the GGPPS enzyme and screening for mutations within the enzyme that led to increased diterpenoid production. In some embodiments, the GGPPS enzyme has one or more of the follow mutations: A162V, G140C, L182M, F218Y, D160G, C184S, K367R, A151T, M185I, D264Y, E368D, C184R, L331I, G262V, R365S, A114D, S239C, G295D, I276V, K343N, P183S, I172T, D267G, I149V, T234I, E153D and T259A.

In some embodiments, the GGPPS enzyme has a mutation in residue S239 and/or residue G295. In certain embodiments, the GGPPS enzyme has the mutation S239C and/or G295D. Mutations in GGPPS that had beneficial effects on diterpenoid production were frequently found to be located between two highly conserved aspartate-rich domains: DDXXXXD and DDXXD (Fig. 4c). A structural analysis of *E. coli* FPP synthase suggested that the two aspartate-rich regions bind three Mg^{2+} ions to facilitate the anchoring of the diphosphate groups of the IPP and DMAPP substrates[44]. Without wishing to be bound by any theory, due to the close proximity to the aspartate motifs and G295 replacement with aspartate, the S239 and/or G295 mutations may affect GGPPS catalysis by improving the binding efficiency of the magnesium ions needed for substrate anchoring.

Combination of a mutant LPS enzyme and a mutant GGPPS enzyme can be expressed in a cell to provide increased production of diterpenoid. In some embodiments, the cell expresses an LPS enzyme containing the mutations M593I and/or Y700F, and a GGPPS enzyme containing the mutations S239C and/or G295D. It should be appreciated that the choice of mutations will in some instances depend on the desired end product. For example, some mutations or combinations of mutations may be selected because they lead to an overall increase in diterpenoid production, while other mutations or combinations of mutations may be selected because they lead to an increase production of one or more specific diterpenoids, such as levopimaradiene, relative to production of other diterpenoids. For example, a cell expressing an LPS enzyme containing the mutation M593I and either Y700A or Y700C produced a selectivity for levopimaradiene of approximately 97 %. A cell expressing both an LPS enzyme containing the mutations M593I and Y700F and a GGPPS enzyme containing

the mutations S239C and G295D was found to improve titer of levopimaradiene by approximately 19 fold over wild-type.

In some embodiments, it may be advantageous to use a cell that has been optimized for production of a diterpenoid. For example, in some embodiments, a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway is used, at least in part, to amplify isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), substrates of GGPPS. In some embodiments, overexpression of one or more components of the non-mevalonate (MEP) pathway is achieved by increasing the copy number of one or more components of the non-mevalonate (MEP) pathway. For example, copy numbers of components at rate-limiting steps in the MEP pathway such as (*dxs*, *ispD*, *ispF*, *idi*) can be amplified, such as by additional episomal expression. In some embodiments, screening for mutations in components of the MEP pathway, or components of other pathways, that lead to enhanced production of a diterpenoid may be conducted through a random mutagenesis screen, or through screening of known mutations. In some embodiments, shotgun cloning of genomic fragments could be used to identify genomic regions that lead to an increase in production of a diterpenoid, through screening cells or organisms that have these fragments for increased production of a diterpenoid. In some cases one or more mutations may be combined in the same cell or organism.

In some embodiments, production of a diterpenoid in a cell can be increased through manipulation of enzymes that act in the same pathway as the enzymes associated with the invention. For example, in some embodiments it may be advantageous to increase expression of an enzyme or other factor that acts upstream of a target enzyme such as an enzyme associated with the invention. This could be achieved by over-expressing the upstream factor using any standard method.

A further strategy for optimization of protein expression is to increase expression levels of one or more genes associated with the invention through selection of appropriate promoters and ribosome binding sites. In some embodiments, this may include the selection of high-copy number plasmids, or low or medium-copy number plasmids. The step of transcription termination can also be targeted for regulation of gene expression, through the introduction or elimination of structures such as stem-loops.

The invention also encompasses isolated LPS and GGPPS polypeptides containing mutations in residues described above, and isolated nucleic acid molecules encoding such polypeptides. As used herein, the terms “protein” and “polypeptide” are used

interchangeably and thus the term polypeptide may be used to refer to a full-length polypeptide and may also be used to refer to a fragment of a full-length polypeptide. As used herein with respect to polypeptides, proteins, or fragments thereof, "isolated" means separated from its native environment and present in sufficient quantity to permit its identification or use. Isolated, when referring to a protein or polypeptide, means, for example: (i) selectively produced by expression cloning or (ii) purified as by chromatography or electrophoresis. Isolated proteins or polypeptides may be, but need not be, substantially pure. The term "substantially pure" means that the proteins or polypeptides are essentially free of other substances with which they may be found in production, nature, or *in vivo* systems to an extent practical and appropriate for their intended use. Substantially pure polypeptides may be obtained naturally or produced using methods described herein and may be purified with techniques well known in the art. Because an isolated protein may be admixed with other components in a preparation, the protein may comprise only a small percentage by weight of the preparation. The protein is nonetheless isolated in that it has been separated from the substances with which it may be associated in living systems, i.e. isolated from other proteins.

Isolated LPS polypeptides can contain mutations in one or more of the following residues: M593, C618, L619, A620, L696, Y700, K723, A727, A729, V731, N769, E777, N838, G854 and I855 (See Fig. 2a). Amino acid residue numbers indicated herein for LPS are based on amino acid numbers in the full-length, wild-type *G. biloba* LPS polypeptide (GenBank Accession No. AF331704). One of ordinary skill in the art would understand, based on protein alignments between *G. biloba* LPS and LPS from other species, how to determine equivalent residues in other species. Isolated LPS polypeptides from species other than *G. biloba,* with mutations in residues that are equivalent to the *G. biloba* residues described above, are also encompassed by the invention.

Non-limiting examples of isolated *G. biloba* LPS polypeptides that can be used alone or in combination in methods associated with the invention include isolated LPS polypeptides that contain one or more of the following mutations: M593I, M593L, M593C, M593S, M593T C618N, L619F, A620T, L696Q, Y700H, Y700F, Y700M, Y700W, K723S, A727S, A729G, V731L, N769A, E777A, N838E, G854T and I855L. Isolated LPS polypeptides from species other than *G. biloba,* with equivalent mutations are also encompassed by the invention.

In some embodiments, the isolated LPS polypeptide contains a mutation in residue M593, alone or in combination with one or more other mutations. For example, the mutation can be M593I or a substitution with another hydrophobic residue such as leucine (M593L). In certain embodiments, the mutation in M593 can be M593C, M593S or M593T. In some embodiments, the isolated LPS polypeptide contains a mutation in residue Y700, alone or in combination with one or more other mutations. For example, the mutation can be Y700H, Y700F, Y700M or Y700W. In some embodiments, the isolated LPS polypeptide contains a mutation in residue A620, alone or in combination with one or more other mutations. In some embodiments, the mutation involves a substitution with a residue that is small and/or hydrophilic. In certain embodiments, the mutation can be A620C, A620G, A620S or A620T. The isolated LPS polypeptide can contain one mutation or multiple mutations. In some embodiments, the isolated LPS polypeptide contains a mutation in M593 and a mutation in Y700. For example the isolated LPS polypeptide can contain the following combinations of mutations: M593I and Y700F, M593I and Y700A, or M593I and Y700C. The isolated LPS polypeptide containing these mutations can also contain one or more other mutations.

Isolated GGPPS polypeptides can contain mutations in one or more of the following residues: A162, G140, L182, F218, D160, C184, K367, A151, M185, D264, E368, C184, L331, G262, R365, A114, S239, G295, I276, K343, P183, I172, D267, I149, T234, E153 and T259. Amino acid residue numbers indicated herein for GGPPS are based on amino acid numbers in the full-length, wild-type *T. canadensis* GGPPS polypeptide (GenBank Accession No. AF081514). One of ordinary skill in the art would understand, based on protein alignments between *T. canadensis* GGPPS and GGPPS from other species, how to determine equivalent residues in other species. Isolated GGPPS polypeptides from species other than *T. canadensis,* with mutations in residues that are equivalent to the *T. canadensis* residues described above, are also encompassed by the invention.

Non-limiting examples of isolated *T. canadensis* GGPPS polypeptides that can be used alone or in combination in methods associated with the invention include isolated GGPPS polypeptides that contain one or more of the following mutations: A162V, G140C, L182M, F218Y, D160G, C184S, K367R, A151T, M185I, D264Y, E368D, C184R, L331I, G262V, R365S, A114D, S239C, G295D, I276V, K343N, P183S, I172T, D267G, I149V, T234I, E153D and T259A. Isolated GGPPS polypeptides from species other than *T. canadensis,* with equivalent mutations are also encompassed by the invention.

In some embodiments, the isolated GGPPS polypeptide contains a mutation in residue S239 and/or residue G295. In certain embodiments, the isolated GGPPS polypeptide has the mutation S239C and/or G295D. The isolated LPS polypeptide containing these mutations can also contain one or more other mutations.

The invention also encompasses nucleic acids that encode for any of the polypeptides described herein, libraries that contain any of the nucleic acids and/or polypeptides described herein, and compositions that contain any of the nucleic acids and/or polypeptides described herein. It should be appreciated that libraries containing nucleic acids or proteins can be generated using methods known in the art. A library containing nucleic acids can contain fragments of genes and/or full-length genes and can contain wild-type sequences and mutated sequences. A library containing proteins can contain fragments of proteins and/or full length proteins and can contain wild-type sequences and mutated sequences. It should be appreciated that the invention encompasses codon-optimized forms of any of the nucleic acid and protein sequences described herein.

The invention encompasses any type of cell that recombinantly expresses genes associated with the invention, including prokaryotic and eukaryotic cells. In some embodiments the cell is a bacterial cell, such as *Escherichia* spp., *Streptomyces* spp., *Zymonas* spp., *Acetobacter* spp., *Citrobacter* spp., *Synechocystis* spp., *Rhizobium* spp., *Clostridium* spp., *Corynebacterium* spp., *Streptococcus* spp., *Xanthomonas* spp., *Lactobacillus* spp., *Lactococcus* spp., *Bacillus* spp., *Alcaligenes* spp., *Pseudomonas* spp., *Aeromonas* spp., *Azotobacter* spp., *Comamonas* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Gluconobacter* spp., *Ralstonia* spp., *Acidithiobacillus* spp., *Microlunatus* spp., *Geobacter* spp., *Geobacillus* spp., *Arthrobacter* spp., *Flavobacterium* spp., *Serratia* spp., *Saccharopolyspora* spp., *Thermus* spp., *Stenotrophomonas* spp., *Chromobacterium* spp., *Sinorhizobium* spp., *Saccharopolyspora* spp., *Agrobacterium* spp. and *Pantoea* spp. The bacterial cell can be a Gram-negative cell such as an *Escherichia coli* (*E. coli*) cell, or a Gram-positive cell such as a species of *Bacillus*. In other embodiments, the cell is a fungal cell such as a yeast cell, e.g., *Saccharomyces* spp., *Schizosaccharomyces* spp., *Pichia* spp., *Paffia* spp., *Kluyveromyces* spp., *Candida* spp., *Talaromyces* spp., *Brettanomyces* spp., *Pachysolen* spp., *Debaryomyces* spp., *Yarrowia* spp. and industrial polyploid yeast strains. Preferably the yeast strain is a *S. cerevisiae* strain. Other examples of fungi include *Aspergillus* spp., *Pennicilium* spp., *Fusarium* spp., *Rhizopus* spp., *Acremonium* spp., *Neurospora* spp., *Sordaria* spp., *Magnaporthe* spp., *Allomyces* spp., *Ustilago* spp., *Botrytis*

spp., and *Trichoderma* spp. In other embodiments, the cell is an algal cell, or a plant cell. It should be appreciated that some cells compatible with the invention may express an endogenous copy of one or more of the genes associated with the invention as well as a recombinant copy. In some embodiments, if a cell has an endogenous copy of one or more of the genes associated with the invention then the methods will not necessarily require adding a recombinant copy of the gene(s) that are endogenously expressed. In some embodiments the cell may endogenously express one or more enzymes from the pathways described herein and may recombinantly express one or more other enzymes from the pathways described herein for efficient production of a terpenoid, such as a diterpenoid.

In some embodiments, one or more of the genes associated with the invention is expressed in a recombinant expression vector. As used herein, a "vector" may be any of a number of nucleic acids into which a desired sequence or sequences may be inserted by restriction and ligation for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA, although RNA vectors are also available. Vectors include, but are not limited to: plasmids, fosmids, phagemids, virus genomes and artificial chromosomes.

A cloning vector is one which is able to replicate autonomously or integrated in the genome in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in the host cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host cell such as a host bacterium or just a single time per host before the host reproduces by mitosis. In the case of phage, replication may occur actively during a lytic phase or passively during a lysogenic phase.

An expression vector is one into which a desired DNA sequence may be inserted by restriction and ligation such that it is operably joined to regulatory sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β-galactosidase, luciferase or alkaline phosphatase), and genes which visibly affect the

phenotype of transformed or transfected cells, hosts, colonies or plaques (e.g., green fluorescent protein). Preferred vectors are those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

As used herein, a coding sequence and regulatory sequences are said to be "operably" joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of that DNA sequence such that the resulting transcript can be translated into the desired protein or polypeptide.

When the nucleic acid molecule that encodes any of the enzymes of the claimed invention is expressed in a cell, a variety of transcription control sequences (e.g., promoter/enhancer sequences) can be used to direct its expression. The promoter can be a native promoter, i.e., the promoter of the gene in its endogenous context, which provides normal regulation of expression of the gene. In some embodiments the promoter can be constitutive, i.e., the promoter is unregulated allowing for continual transcription of its associated gene. A variety of conditional promoters also can be used, such as promoters controlled by the presence or absence of a molecule.

The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. In particular, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. The vectors of the invention may optionally include 5' leader or signal sequences.

The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA (RNA). That heterologous DNA (RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell. Heterologous expression of genes associated with the invention, for production of a terpenoid, such as a diterpenoid, is demonstrated in Example 1 using *E. coli*. The novel method for producing diterpenoids can also be expressed in other bacterial cells, fungi (including yeast cells), plant cells, etc.

A nucleic acid molecule that encodes the enzyme of the claimed invention can be introduced into a cell or cells using methods and techniques that are standard in the art. For example, nucleic acid molecules can be introduced by standard protocols such as transformation including chemical transformation and electroporation, transduction, particle bombardment, etc. Expressing the nucleic acid molecule encoding the enzymes of the claimed invention also may be accomplished by integrating the nucleic acid molecule into the genome.

In some embodiments one or more genes associated with the invention is expressed recombinantly in a bacterial cell. Bacterial cells according to the invention can be cultured in media of any type (rich or minimal) and any composition. As would be understood by one of ordinary skill in the art, routine optimization would allow for use of a variety of types of media. The selected medium can be supplemented with various additional components. Some non-limiting examples of supplemental components include glucose, antibiotics, IPTG for gene induction, ATCC Trace Mineral Supplement, and glycolate. Similarly, other aspects of the medium, and growth conditions of the cells of the invention may be optimized through routine experimentation. For example, pH and temperature are non-limiting examples of factors which can be optimized. In some embodiments, factors such as choice of media, media supplements, and temperature can influence production levels of terpenoids, such as diterpenoids. In some embodiments the concentration and amount of a supplemental component may be optimized. In some embodiments, how often the media is supplemented

with one or more supplemental components, and the amount of time that the media is cultured before harvesting a terpenoid, such as a diterpenoid, is optimized.

According to aspects of the invention, high titers of a diterpenoid such as levopimaradiene, are produced through the recombinant expression of genes associated with the invention, in a cell. As used herein "high titer" refers to a titer in the milligrams per liter (mg L^{-1}) scale. The titer produced for a given product will be influenced by multiple factors including choice of media. In some embodiments the total diterpenoid titer is at least 10 mg L^{-1}. For example the titer may be 10, 20, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900 or more than 900 mg L^{-1} including any intermediate values. In some embodiments, a cell that expresses an LPS enzyme containing the mutations M593I and Y700F, and a GGPPS enzyme containing the mutations S239C and G295D can produce a total diterpenoid titer of approximately 800 mg L^{-1} in approximately 168 hours.

The liquid cultures used to grow cells associated with the invention can be housed in any of the culture vessels known and used in the art. In some embodiments large scale production in an aerated reaction vessel such as a stirred tank reactor can be used to produce large quantities of terpenoids, such as diterpenoids, that can be recovered from the cell culture. In some embodiments, the terpenoid is recovered from the gas phase of the cell culture, for example by adding an organic layer such as dodecane to the cell culture and recovering the terpenoid from the organic layer.

Diterpenoids, such as levopimaradiene, produced through methods described herein have widespread applications. Levopimaradiene is a key diterpenoid precursor of ginkgolides which can be administered for a variety of therapeutic purposes including improving vascular function, inhibiting thrombosis and embolism, neuroprotective functions, and cancer treatment. Terpenoid pathways also lead to compounds used in flavors, cosmetics, and biofuels. Furthermore, methods described herein to search for mutations in LPS can be applied to other diterpenoid synthases such as taxadiene synthase. GGPPS mutations described herein can also be applied to synthesis of precursors for other plant diterpenoids including cancer therapeutics such as Taxol, plant growth hormones such as gibberellins and food products such as the natural sweetener steviol glycoside.

Examples

Example 1: Harnessing the evolvability of a terpenoid biosynthetic pathway for overproduction and selectivity control.

Introduction

The engineering of secondary metabolite biosynthesis in heterologous microorganisms is a promising approach to produce drug precursors in a scalable manner. However, secondary metabolite pathways are typically low-yielding and produce side products. Herein, these limitations were addressed by harnessing the evolvability of a plant-derived terpenoid pathway to efficiently synthesize levopimaradiene, the gateway precursor of the bioactive ginkgolides. Variants of geranylgeranyl diphosphate synthase and levopimaradiene synthase were created to uncover mutations that confer divergent phenotypes in *Escherichia coli*. Sequence space explorations by random and rational mutagenesis identified combinations of mutations that increased levopimaradiene synthesis up to 19-fold over the wild-type pathway, and reduced the abietadiene and sandaracopimaradiene isomers. In bench-scale controlled culture conditions, strains harboring the highest-producing pathway variant resulted in ~700 mg/L levopimaradiene. This pathway reprogramming framework should expedite engineered biosynthesis applications for large-scale pharmaceutical production, and facilitate the overproduction of other chemicals hitherto only derived from natural resources.

Results

Probing the LPS putative binding pocket by phylogenetic-based mutations

The simultaneous expression of the wild-type GGPPS and LPS in a pre-engineered *E. coli* strain overexpressing the MEP pathway resulted in the production of ~27 mg/L diterpenoid mixture in a 2-mL culture. In this product mixture, levopimaradiene (87%) was accompanied by abietadiene (11%), sandaracopimaradiene (2%), and neoabietadiene (trace amounts), as identified by gas chromatography – mass spectroscopy (GC-MS) (Fig. 1b, Fig. 6b). This wild-type phenotype provided the baseline comparison used in the selection of evolved pathway variants. The principal challenge in exploring the evolvability of LPS through the generation of large mutant libraries is the lack of a suitable high-throughput screen. Therefore, a structure-guided method was implemented herein to allow the identification of tunable residues within LPS. The crystal structure of LPS is currently not

available. However, because the tertiary folds of terpene cyclases are similar, the only available structure of one such enzyme, that of 5-epi-aristolochene synthase[30] (EAS), was used to thread the putative second active site of LPS.

The second active site was focused on because mutations in this site within other terpene cylases impacted their 'plasticity'[26, 31-33]. In the second active site of an "LPS-type" enzyme, the bicylic (+)-copalyl diphosphate (CPP) intermediate (derived from the deprotonation of GGPP in the first active site) undergoes a diphosphate-ionization cyclization. The resulting C8-sandaracopimarenyl cation intermediate is further deprotonated at two alternative sites to release isopimaradiene or sandaracopimaradiene end products. However, this intermediate can also undergo intramolecular proton transfer and 1,2-methyl migration to yield abietenyl cation. Subsequent deprotonation of abietenyl cation at four possible sites then produce abietadiene, levopimaradiene, neoabietadiene, and palustradiene[28, 29].

To allow sufficient sampling of the three dimensional space, fifteen residues within the 10 Å solvation layer of the LPS model were probed (Fig. 2a). Although residues far from the binding pocket may contribute to enzyme evolvability, consideration of these residues increases the library size to a degree that is experimentally intractable. Amino acid substitutions within related enzymes (phylogenetically-based mutation)[26, 34] were used to perturb the selected fifteen residues of LPS. These residues were replaced with those derived from paralogous "LPS-type" enzymes that are functionally different from LPS (Fig. 2b), namely *Abies grandis* abietadiene synthase (AS), *Picea abies* AS, and *P. abies* isopimardiene synthase (ISO). In AS, abietadiene, levopimaradiene, and neoabietadiene are synthesized in almost equal proportion[32]. On the other hand, ISO produces only isopimaradiene[31]. Mutations of M593I, C618N, L619F, A620T, L696Q, K723S, A729G, N838E, G854T, and I855L were created based on residues in *A. grandis* and *P. abies* AS, whereas Y700H, A727S and V731L were created based on *P. abies* ISO. Alanine was used to replace Asn769 and Glu777 because these amino acids are conserved throughout LPS, AS, and ISO (Fig. 2c).

The pre-engineered *E. coli* expressing the wild-type GGPPS provided an *in vivo* screening system for titer and product distribution changes by the LPS mutations. The profiles of diterpenoid product distribution resulting from expressing LPS mutants M593I, C618N, L619F, A620T, L696Q, K723S, V731L, N838E, G854T, and I855L were observed to be similar to expression of wild type LPS (Fig. 2c). Hence, with respect to LPS product selectivity, these mutations were rendered neutral or close to neutral. Diterpenoid

productivities resulting from expressing LPS mutants L619F, A620T and G854T were also not significantly changed compared to wild-type LPS (within 50%). However, diterpenoid production levels were notably altered by expressing mutants M593I, C618N, L696Q, K723S, V731L, N838E, and I855L (Fig. 2c). The highest total diterpenoid production increase (~3.7-fold) was mediated by expressing LPS M593I. In all cases, expression of these mutants did not significantly affect product distribution. Expression of LPS mutant Y700H, however, resulted in a significant alteration of diterpenoid product distribution by abolishing abietadiene synthesis and increasing sandaracopimaradiene proportion (Fig. 2c). Although a single mutation in *P. abies* AS, corresponding to Tyr700, did not result in product selectivity changes *in vitro*, it promoted product selectivity shift when combined with other mutations[31]; hence Tyr700 may play an important role in mediating the evolvability of "LPS-type" enzymes. Additionally, the expression of the A729G mutant resulted in the exclusive production of sandaracopimaradiene; however, it was concomitant with a reduction of productivity by ~98% (Fig. 2). Finally, diterpenoid production was not observed in systems expressing LPS A727S, N769A, and E777A variants. These residues fell within ~4.7 Å from substrate in the homology model; therefore, it was not surprising that the mutations were deleterious to LPS activity given the close proximity to the substrate.

Mutational enrichment of tunable LPS residues

The previous results pointed to mutations in LPS that significantly affected production phenotype, namely M593I and Y700H. Although the preliminary mutation of Ala729 imparted product selectivity changes, it was excluded from further analysis because even a conservative replacement such as glycine was deleterious. From analyzing the structural model, Met593 was observed to be located at the posterior of the binding pocket, whereas Tyr700 is positioned at the entrance (in close vicinity of the DDXXD magnesium binding motif). To obtain the complete LPS evolvability profile by these residues, all amino acids were sampled through saturation mutagenesis. Additionally, the effects of expressing the saturation mutagenesis library of Ala620 was explored because a mutation at this position in *A. grandis* AS changed its product selectivity *in vitro*[32].

From the saturation mutagenesis library of Met593, two substitutions were found that conferred significant productivity improvement (Fig. 3a). In addition to isoleucine which was discovered in the phylogenetic-based mutation, the replacement with another hydrophobic residue similar in size as methionine, i.e. leucine, increased diterpenoid productivity by ~2-

fold without significantly changing product distribution (Fig. 3a). Based on the structural significance of this position, this productivity improvement appeared to be caused by the disruption of H-bonding at the end of the binding pocket, thus increasing the flexibility of the cavity to better fit the CPP substrate. Therefore, the M593I mutation likely resulted in the highest production increase (~3.7-fold) because isoleucine is the most hydrophobic amino acid[35]. Substitutions with much smaller residues than methionine only yielded moderate production improvement (less than 2-fold in the case of cysteine, serine, and threonine), and were disruptive in the case of alanine, glycine, and valine. Furthermore, substitutions with amino acids longer than five heavy atoms and those with bulky rings such as phenylalanine, tyrosine, and tryptophan also consistently decreased or abolished activity. These trends are consistent with the requirement for the substrate to have an unobstructed cyclization pocket, as the center of the bend is proximal to this residue. Moreover, replacements with hydrophilic amino acids[36] such as aspartic acid, glutamic acid, lysine, and arginine also generally reduced productivity because their ability to form their own H-bonding may reduce the capacity of the binding pocket.

The replacement of Tyr700 with phenylalanine, methionine, and tryptophan improved productivity up-to ~5-fold (Fig. 3b). The reaction cascade toward the formation of abietenyl cation requires an energetically unfavorable transition from a tertiary to a secondary carbocation[29]. Therefore, it was postulated that the latter species is stabilized by the ionic interaction with the paired diphosphate anion that is chelated by the magnesium ion[32]. Tyr700 is located within close proximity to the magnesium binding site, thus the absence of the hydroxyl group in amino acids that are similar to tyrosine may allow the repositioning of the magnesium closer to the aspartate-rich region; hence, increasing reaction efficiency by improving the chelation of the diphosphate group. A few mutations, i.e. replacements with aspartic acid, histidine, proline, arginine, and lysine, abolished abietadiene synthesis in the product mixture, and conferred a decrease in productivity (Fig. 3b). The replacement with positively charged residues or a helix breaker (proline) might cause a misalignment of the diphosphate anion that impaired catalysis or prevented the deprotonation of abietenyl cation at carbon position f (Fig. 1b) to create abietadiene.

Finally, the sampling of all amino acid substitutions of Ala620 revealed that only replacement with residues similar to alanine (small or hydrophilic) (cysteine, glycine, serine, and threonine) as well as valine retained LPS activity; whereas other substitutions were destructive or deleterious (Fig. 3c). A few destructive mutations (replacements with aspartic

acid, leucine, asparagine) also destabilized abietenyl deprotonation to yield abietadiene. Therefore, Ala620 in LPS did not appear to control product selectivity and productivity in LPS, yet it was important for catalysis.

Combinatorial LPS mutations

In laboratory experiments, the beneficial effect of single mutations are often additive[33, 37, 38]. Therefore, the production improvement resulting from expressing the LPS M593I variant encouraged investigation of the effect of this beneficial mutation in combination with saturation mutagenesis of Tyr700. As shown in Fig. 3d, this combination successfully integrated the advantageous properties of the individual mutations, resulting in one variant (M593I/Y700F) that increased diterpenoid titer by 10-fold over the expression of the wild-type enzyme without significant changes in product distribution. Interestingly, the expression of most LPS variants carrying the double mutations resulted in the reduction of abietadiene proportion (Fig. 3d). In two variants, M593I/Y700A and M593I/Y700C, the reduction of abietadiene was complemented by an increase in levopimaradiene, resulting in mixtures that contained up to 97% levopimaradiene. The combinations of M593I with Y700A and Y700C also conferred productivity improvement by ~6.3- and ~5.8-fold, respectively. Thus, it appeared that these mutations facilitated the stabilization of the intermediates that led to favoring abietenyl intermediate formation yet prevented proton extraction from route f (Fig. 1b). It is noteworthy that using the M593I mutant LPS as a parent for the subsequent saturation-mutagenesis only conferred three variants that decreased or abolished diterpenoid productivity (Fig. 3d). In contrast, previous saturation mutagenesis of Tyr700 resulted in five pathway variants that decreased/abolished diterpenoid production (Fig. 3b). Although a single M593I mutation was rendered neutral with respect to product selectivity, this mutation restored LPS activity upon subsequent defective mutations by aspartic acid and histidine replacement at residue 700. Thus M593I appeared to facilitate subsequent mutational robustness, a paradigm of neutral genetic drift[39].

Random mutagenesis of GGPPS

The generation of a high-producing pathway was extended by the creation of a GGPPS library. As an up-stream enzyme of LPS, GGPPS catalyzes the formation of the linear polyprenyl (C_{20}) diphosphate starter unit by the sequential elongation of IPP with the allylic monomer. Concomitant with diterpenoid production increase, methyl jasmonate

elicitation in Taxus cell culture elevated GGPPS expression level together with the respective downstream cyclase[40]. These results suggested that together with the cyclase, GGPPS is an important target in the diterpenoid pathway for increasing productivity. To optimize levopimaradiene production, *T. canadensis* GGPPS was incorporated into the pathway assembly because this enzyme has high specificity toward FPP to synthesize GGPP.

Although the structural information of a plant GGPPS from an angiosperm origin is available[41], the crystal structure for a gymnosperm GGPPS has not been solved. Furthermore, the folding similarity of gymnosperm GGPPS enzymes and their angiosperm analogs are not known. Despite catalyzing essentially the same enzymatic reaction, GGPPS enzymes are known to exhibit wide structural diversity among organisms[41]. Therefore, based on secondary structure analysis[42], the notable division of gymnosperm from angiosperm GGPPS enzymes may imply significant tertiary fold differences. The lack of a suitable structural guide prompted us to devise a stochastic mutational approach to evolve *T. canadensis* GGPPS. To enable a facile high-throughput screening method for isolating improved GGPPS variants, we utilized a lycopene biosynthetic pathway consisting of *crt*B and *crt*I as a colorimetric reporter (Fig. 4a). In this system, the expression of wild-type GGPPS resulted in colonies with orange coloration. Improved GGPPS variants from the mutagenesis were identified by the improvement of lycopene production in the cell, as determined by red coloration. Fifteen *ggpps* variants were isolated from colonies exhibiting red coloration. To assess the potential for improving levopimaradiene production *in vivo*, the fifteen mutant GGPPS isolates were co-expressed with the high-producing LPS M593I/Y700F mutant in the pre-engineered *E. coli* strain. Five GGPPS variants did not confer a levopimaradiene increase indicating false positives obtained from the colorimetric screening. However, the co-expression of ten GGPPS mutants resulted in diterpenoid production improvement (Fig. 4b.). The expression of mutant G10 resulted in the highest diterpenoid production increase (~1.7-fold) over the pathway harboring the wild-type GGPPS and the LPS M593I/Y700F, representing a ~20-fold total increase (~19-fold levopimaradiene increase) over the pathway harboring wild-type GGPPS and LPS (Fig. 4b, Table 6).

Sequence analysis of G10 revealed that two positions were mutated, namely S239C and G295D (Fig. 4c). Amino acid alignment with GGPPS sequences from other plants[43] showed that most beneficial mutations are located in the region in between the two highly conserved aspartate-rich DDXXXXD and DDXXD domains (Fig. 4c). A structural analysis of *E. coli* FPP synthase suggested that the two aspartate-rich regions bound three Mg^{2+} ions

to facilitate the anchoring of the diphosphate groups of the IPP and DMAPP substrates[44]. Therefore due to the close proximity to the aspartate motifs and Gly295 replacement with aspartate, the mutations in G10 may affect GGPPS catalysis by improving the binding efficiency of the magnesium ions needed for substrate anchoring.

Levopimaradiene overproduction in controlled culture conditions

The performance of the pre-engineered *E. coli* strain expressing the highest-producing levopimaradiene pathway (consisting of GGPPS S239C/G295D and LPS M593I/Y700F) was assessed in small bioreactors (1 L) under controlled conditions. The total diterpenoid titer reached a maximum of ~800 mg/L in 168 h, and levopimaradiene constituted ~700 mg/L of total diterpenoids (Fig. 5a). Using this engineered strain, 10 g/L glycerol was almost depleted after 56 h. Therefore 3 g/L glycerol was added to the culture every 8 h after this time point (Fig. 5b). Despite the relatively rapid consumption of glycerol, acetate only accumulated below 1 g/L throughout the cultivation. This suggested a significant diversion of flux from acetyl-CoA (acetate precursor) because the up-stream precursors, pyruvate and glyceraldehyde 3-phosphate were efficiently channeled by the engineered pathway for synthesizing the diterpenoid products. Overall, this experiment demonstrated that the production improvement obtained from the new pathway translated well toward a larger cultivation.

Discussion

Herein, a combination of rational and random mutational searches were used to uncover cryptic genetic variations in an engineered plant pathway that imparted levopimaradiene production changes in *E. coli*. Overall, the identification of as few as one to two mutations of LPS and GGPPS to generate changes in diterpenoid production or product distribution highlights the susceptibility of promiscuous secondary metabolic enzymes to new functions. Structure-guided analysis followed by saturation mutagenesis revealed several mutations in LPS that conferred diterpenoid production improvements and product selectivity changes. Notably, when M593I and Y700F mutations were combined, an additional ~6.5-fold production increase was achieved over the M593I mutation alone (Fig. 3a, 3d, and Tables 2 and 5). This result presents another example of the additivity of beneficial mutations found in laboratory evolution[33, 37, 38]. Furthermore, the combination of M593I either with Y700A or Y700C also reduced the proportion of undesired isomers, increasing selectivity for

levopimaradiene up to 97%, while still maintaining improved productivity (Fig. 3d, and Table 5). In general, the strategies to search for beneficial mutations in LPS may also be applied to other important diterpenoid cyclases, such as the taxadiene synthase, because terpene cyclases have similar tertiary folds[45]. Several GGPPS variants that conferred diterpenoid production increase were identified by random mutagenesis and utilizing a lycopene pathway as a screening system. A GGPPS variant harboring S239C/G295D, when combined with LPS M593I/Y700F created a mutant pathway that improved the levopimaradiene titer by ~19-fold over wild-type. In controlled culture conditions, diterpenoid titer from the pre-engineered *E. coli* harboring the highest-producing mutant pathway reached ~800 mg/L, of which ~700 mg/L constituted levopimaradiene. Readily available levopimaradiene opens the possibility for utilization as a biotranformation substrate in plant tissue/cell culture to synthesize ginkgolides[7] or derivatization for developing new pharmaceuticals using synthetic chemistry[46]. Additionally, because GGPPS is required in all diterpenoid pathways, the GGPPS mutant identified in this work should also find application for the synthesis of other precursors for other important plant diterpenoids such as Taxol (cancer chemotherapeutics), gibberellins (plant growth hormones), and steviol glycosides (a natural sweetener that does not induce glycemic responses).

The approval of more than 100 new natural product-derived drugs for clinical trial in 2007 signifies the long-standing role of these molecules as effective therapeutics. Yet, this figure represents about a 30% drop since 2001[47]. One of the major challenges in many natural product research efforts is the reliance on bioprospecting, which typically generates low yield. This work demonstrated that by transferring and reengineering a heterologous biosynthetic pathway, the high level production of a plant-derived pharmaceutical can be achieved in a microbial host. This pathway 'reprogramming' framework should further enhance the extent of production improvement via metabolic engineering and complement a recently developed tool to mediate metabolite channeling *in vivo*[48]. In a broader sense, because terpenoid pathways also lead to compounds used in flavors, cosmetics, and biofuels, this strategy should also be readily extended to overproduce many commercially important compounds using microbial biotechnology.

Methods

Cloning and pathway construction

The sequences of *ggpps*[43] and *lps*[49] were obtained from *Taxus canadensis* and *G. biloba*, respectively (Genbank accession codes: AF081514 and AF331704). Genes were custom-synthesized (DNA 2.0) to incorporate *E. coli* codon bias, remove restriction sites for cloning purposes, and establish a ~50% GC-content. Nucleotides corresponding to the 98 N-terminal amino acids of GGPPS (plastid transit peptide) were removed by designing custom oligonucleotides to generate mature proteins as previously described[43]. In the case of LPS, truncation of 40 N-terminal amino acids was chosen because its incorporation into the levopimaradiene pathway gave rise to the most stable diterpenoid production in comparison to 60- and 80-amino acid truncation. In all cases, a start codon was introduced in the truncated gene fragments. For creating mutagenesis templates and sequencing purposes, *ggpps* and *lps* were individually cloned into pTrc99A (GE Healthcare) into the *Hind*III–*EcoR*I and *EcoR*I–*Sal*I restriction sites, respectively.

The levopimaradiene pathways (wild type and mutants) were constructed by cloning PCR fragments of *ggpps* and *lps* into the *Hind*III–*EcoR*I and *EcoR*I–*Sal*I sites of pTrcMod[50] to create pTrcGGPPS-LPS. To allow high throughput screening of GGPPS mutants, the biosynthetic gene cluster consisting of *crt*B and *crt*I derived from plasmid pAC-LYC[16] were cloned into the *EcoR*I – *Sal*I sites of pTrcMod to yield pTrcCRT. The mutant *ggpps* library was subsequently cloned into pTrcCRT in between the *Hind*III and *EcoR*I sites to create pTrcGGPPS*-CRT. In all cases, *E. coli* MG1655 Δ (*end*A, *rec*A) overexpressing the MEP pathway was used as the expression strain of the various pathways (wild-type and mutant levopimaradiene pathways, wild-type and mutant lycopene pathways). The episomal overexpression of the MEP pathway was mediated by first cloning the operon consisting of *dxs*, *idi*, and *ispFD* into the *Nco*I–*Kpn*I of pTrcMod to yield pTrcMEP. The *trc* promoter and *lacI*q sequences were then amplified together with the MEP operon and sub-cloned into the *Pme*I and *Mlu*I sites of pACYC184 to create plasmid pACMEP.

Culture growth and library analysis

Single transformants of pre-engineered *E. coli* strains harboring pACME or their mutant variants were cultivated for 18 h at 30°C in Luria-Bertani (LB) medium. For library characterization, these preinnocula were used to seed fresh 2-mL cultures at a starting A_{600} of 0.1. The medium was composed of yeast extract, 5 g/L; Trypton, 10 g/L; glycerol, 15 g/L; NaCl, 10 g/L; HEPES, 100 mM; pH was adjusted to 7.6. Cultures were grown for 120 h at 22°C prior to diterpenoid analysis. Scale-up experiments were done in 1-L bioreactors using.

The media composition was as follow: KH_2PO_4, 13.3 g/L; $(NH_4)_2HPO_4$, 4 g/L; citric acid, 1.7 g/L; EDTA, 0.0084 g/L; $CoCl_2$, 0.0025 g/L; $MnCl_2$, 0.015 g/L; $CuCl_2$, 0.0015 g/L; H_3BO_3, 0.003 g/L; Na_2MoO_4, 0.0025 g/L; $Zn(CH_3COO)_2$, 0.008 g/L; Fe(III) citrate, 0.06 g/L; thiamine, 0.0045 g/L; $MgSO_4$, 1.3 g/L; yeast extract, 5 g/L; antifoam B, 3 mL/L; pH was maintained at 7.0. Glycerol was initially supplied at 10 g/L, it was intermittently fed so that the concentration did not reach below 3 g/L. The aeration level was set to 0.5 vvm, dissolved oxygen level was controlled at more than 20% during the course of fermentation by increasing agitation speed. All cultures were supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. To minimize the loss of diterpenoids due to air-stripping, 2% dodecane was added into the culture.

For analysis of small-scale cultivations (libraries), 1 mL hexane was added into 1.5 mL culture aliquots and vortexed for 30 min. The mixture was centrifuged to separate the organic layer. For analysis of bioreactor cultivations, 1 μL of the dodecane layer was diluted to 200 μL with hexane. In both cases, 1 μL of hexane (containing the analytes) was analyzed by GC-MS (Varian Saturn 3800 GC attached to a Varian 2000 MS). The sample was injected into a HP5ms column 30m x 250 μM x 0.25 μM thickness (Agilent). Helium (ultra purity) at a flow rate 1.0 ml/min was used as a carrier gas. The oven temperature was first kept constant at 50° C for 1 min, and then increased to 220° C at the increment of 10° C/min, and finally held at this temperature for 10 min. The injector and transfer line temperatures were set at 200° C and 250°C, respectively. Because levopimaradiene, abietadiene, and sandaracopimaradiene are not commercially available, taxadiene, a diterpenoid possessing the same molecular mass as levopimaradiene, abietadiene, sandaracopimaradiene was used to construct a calibration curve for the peak areas obtained from the GC-MS.

Molecular modeling

The 3D structural model of LPS was built based on EAS (Protein Data Bank ID code 5EAT). Sequence alignment (Fig. 7) was performed with the clustalW method with standard gap penalties. While LPS contains 323 residues in excess of EAS, they aligned almost exclusively at the proximity of the second active side (towards the C-terminus), with a virtually gapless alignment. The CHARMM molecular modeling software with the CHARMm27 parameter set was used to mutate residues. Partial atomic charges needed for the substrate were obtained quantum mechanically with the Gaussian program using the 6-31G* basis set.

Mutant library generation and screening

The introduction of point mutations and saturation mutagenesis in *lps* were performed using QuikChange II XL (Stratagene). Nucleotide changes were set by custom designed oligonucleotides (Table 7). Subsequent to sequencing to verify nucleotide changes, the *lps* variants were used to replace the wild-type *lps* in pTrcGGPPS-LPS and subjected to expression in the pre-engineered *E. coli* for production analysis. Ramdom mutagenesis library of *ggpps* was created by error-prone (EP) PCR at low mutation rate using GeneMorph II (Stratagene). A pool of plasmid pTrcGGPPS*-CRT was isolated from more than ~10^6 transformants of *E. coli* DH10B. The plasmid library was then used to transform the *E. coli* strain overexpressing the MEP pathway for colorimetric screening. Colonies that displayed bright red coloration were isolated after incubation at 25°C for 3 days (as visualized on Luria-Bertani solid medium containing 75 μg/mL ampicillin and 25 μg/mL chloramphenicol). Following plasmid extraction and sequencing, the mutant *ggpps* genes were used as a pool in the next round of EP PCR. As a control, the integration of wild-type *ggpps* into the lycopene pathway gave rise to orange colored transformants. The iteration of mutation and screening was stopped after the 2nd round of mutant collection, as no colony that displayed higher red coloration was identified in the 3rd round of EP PCR.

Table 1: Diterpenoid production from pre-engineered *E. coli* strains harboring GGPPS and LPS mutants (phylogenetically-based mutations)

LPS mutation	Titer (mg/L)
WT	26.5
M593I	98.6
C618N	4.0
L619F	12.7
A620T	33.9
L696Q	42.0
Y700H	11.5
K723S	48.0
A727S	0.0
A729G	0.6
V731L	2.3
N769A	0.0
E777A	0.0
N838E	58.5
G854T	36.8
I855L	17.7

WT, wild type LPS. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 2: Diterpenoid production from pre-engineered *E. coli* strains harboring GGPPS and LPS mutants (saturation mutagenesis of Met593)

LPS Met593 Mutation	Product Selectivity (%)			Titer (mg/L)
	1	2	3	
WT	87	11	2	26.5
Ala	65	26	9	16.2
Cys	65	28	7	48.4
Asp	51	15	34	10.3
Glu	30	7	63	9.7
Phe	ND	ND	100	3.8
Gly	67	23	10	1.3
His	ND	ND	ND	0.0
Ile	84	12	5	98.6
Lys	ND	ND	ND	0.0
Leu	80	13	7	55.2
Asn	75	11	14	40.4
Pro	49	8	43	12.6
Gln	43	TA	57	8.6
Arg	TA	ND	100	4.6
Ser	80	11	9	39.8
Thr	78	16	6	40.2
Val	71	22	6	18.2
Trp	ND	ND	ND	0.0
Tyr	ND	ND	ND	0.0

WT, wild type LPS; TA, trace amounts (< 0.1%); ND, not detected. Levopimaradiene, 1; abietadiene, 2; sandaracopimaradiene, 3. Neoabietadiene is not included in the table because it was only produced in trace amounts in all strains. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 3: Diterpenoid production from pre-engineered *E. coli* strains harboring GGPPS and LPS mutants (saturation mutagenesis of Ala620)

LPS Ala620 Mutation	Product Selectivity (%)			Titer (mg/L)
	1	2	3	
WT	87	11	2	26.5
Cys	87	10	3	40.0
Asp	97	2	1	1.7
Glu	ND	ND	ND	0.0
Phe	ND	ND	ND	0.0
Gly	86	12	2	29.2
His	100	ND	ND	0.9
Ile	ND	ND	ND	0.0
Lys	ND	ND	ND	0.0
Leu	97	2	1	7.3
Met	ND	ND	ND	0.0
Asn	97	2	1	2.6
Pro	ND	ND	ND	0.0
Gln	ND	ND	ND	0.0
Arg	ND	ND	ND	0.0
Ser	92	8	TA	42.2
Thr	87	11	2	33.9
Val	92	8	TA	42.5
Trp	ND	ND	ND	0.0
Tyr	ND	ND	ND	0.0

WT, wild type LPS; TA, trace amounts (< 0.1%); ND, not detected. Levopimaradiene, 1; abietadiene, 2; sandaracopimaradiene, 3. Neoabietadiene is not included in the table because it was only produced in trace amounts in all strains. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 4: Diterpenoid production from pre-engineered *E. coli* strains harboring GGPPS and LPS mutants (saturation mutagenesis of Tyr700)

LPS Tyr700 Mutation	Product Selectivity (%)			Titer (mg/L)
	1	2	3	
WT	87	11	2	26.5
Ala	81	9	10	75.6
Cys	76	7	17	59.1
Asp	81	ND	19	13.3
Glu	64	7	29	46.3
Phe	79	16	5	133.5
Gly	79	6	15	36.1
His	74	ND	26	6.2
Ile	60	6	34	31.5
Lys	TA	ND	100	4.3
Leu	72	7	21	48.2
Met	80	13	7	132.8
Asn	70	9	21	60.2
Pro	56	ND	44	8.3
Gln	59	6	35	41.2
Arg	33	ND	67	5.0
Ser	78	7	15	84.9
Thr	72	7	21	65.4
Val	56	6	38	31.3
Trp	84	6	10	100.7

WT, wild type LPS; TA, trace amounts (< 0.1%); ND, not detected. Levopimaradiene, 1; abietadiene, 2; sandaracopimaradiene, 3. Neoabietadiene is not included in the table because it was only produced in trace amounts in all strains. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 5: Diterpenoid production from pre-engineered *E. coli* strains harboring GGPPS and LPS M593I mutants (saturation mutagenesis of Tyr700)

LPS (M593I) Tyr700 mutation	Product Selectivity (%)			Titer (mg/L)
	1	2	3	
WT	87	11	2	26.5
Ala	91	2	7	167.8
Cys	97	1	2	155.7
Asp	85	TA	15	66.1
Glu	60	TA	40	92.4
Phe	84	9	7	273.8
Gly	80	TA	20	59.2
His	70	TA	30	63.8
Ile	65	TA	35	78.6
Lys	ND	ND	ND	0.0
Leu	73	2	25	97.4
Met	89	2	9	132.1
Asn	79	TA	21	93.7
Pro	0	TA	ND	0.0
Gln	60	TA	40	26.9
Arg	ND	ND	ND	0.0
Ser	84	TA	16	48.1
Thr	73	TA	27	22.7
Val	91	TA	9	83.5
Trp	63	TA	37	51.6

WT, wild-type LPS; TA, trace amounts (< 0.1%); ND, not detected. Levopimaradiene, 1; abietadiene, 2; sandaracopimaradiene, 3. Neoabietadiene is not included in the table because it was only produced in trace amounts in all strains. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 6: Diterpenoid production from pre-engineered *E. coli* strains harboring isolated GGPPS mutants and the LPS M593I/Y700F variant

GGPPS mutation	Titer (mg/L)
WT	273.8
G1	261.3
G2	396.2
G3	343.2
G4	316.8
G5	257.1
G6	242.4
G7	380.9
G8	351.0
G9	350.7
G10	468.7
G11	211.8
G12	366.8
G13	411.1
G14	287.2
G15	406.7

WT, wild type GGPPS. Quantification of production was determined based on sampling an average of three independent *E. coli* colonies harboring the mutated pathway. Standard deviations were lower than 5%.

Table 7: Custom oligonucleotides used for LPS mutagenesis

Mutation	5'-3' Sequence
F-C618N	CGTACGCAAAAACCTCTAACCTGGCCGTAATCCTGG (SEQ ID NO:5)
R-C618N	CCAGGATTACGGCCAGGTTAGAGGTTTTTGCGTACG (SEQ ID NO:6)
F-L619F	GCAAAAACCTCTTGCTTCGCCGTAATCCTGGACGATC (SEQ ID NO:7)
R-L619F	GATCGTCCAGGATTACGGCGAAGCAAGAGGTTTTTGC (SEQ ID NO:8)
F-L696Q	GTAAAGTTTGGGAGGGCCAGCTGGCCTCCTATAC (SEQ ID NO:9)
R-L696Q	GTATAGGAGGCCAGCTGGCCCTCCCAAACTTTAC (SEQ ID NO:10)
F-K723S	GTATGTCGAGAACGCTAGTGTTAGCATCGCGCTGG (SEQ ID NO:11)
R-K723S	CCAGCGCGATGCTAACACTAGCGTTCTCGACATAC (SEQ ID NO:12)
F-A727S	CTAAAGTTAGCATCTCGCTGGCGACCGTTGTTCTG (SEQ ID NO:13)
R-A727S	CAGAACAACGGTCGCCAGCGAGATGCTAACTTTAG (SEQ ID NO:14)
F-A729G	CTAAAGTTAGCATCGCGCTGGGGACCGTTGTTCTG (SEQ ID NO:15)
R-A729G	CAGAACAACGGTCCCCAGCGCGATGCTAACTTTAG (SEQ ID NO:16)
F-V731L	CATCGCGCTGGCGACCCTTGTTCTGAACTC (SEQ ID NO:17)
R-V731L	GAGTTCAGAACAAGGGTCGCCAGCGCGATG (SEQ ID NO:18)
F-N769A	CCGGCCGTCTGATTGCCGACACCAAAACCTATCAG (SEQ ID NO:19)
R-N769A	CTGATAGGTTTTGGTGTCGGCAATCAGACGGCCGG (SEQ ID NO:20)
F-E777A	CCAAAACCTATCAGGCTGCACGTAACCGTGG (SEQ ID NO:21)
R-E777A	CCACGGTTACGTGCAGCCTGATAGGTTTTGG (SEQ ID NO:22)
F-N838E	CGTCGTCTGCTGTTCGAGACCGCGCGTGTAATGC (SEQ ID NO:23)
R-N838E	GCATTACACGCGCGGTCTCGAACAGCAGACGACG (SEQ ID NO:24)
F-G854T	GTACCGCGATGGCTTCACCATCAGCGATAAAGAAATG (SEQ ID NO:25)
R-G854T	CATTTCTTTATCGCTGATGGTGAAGCCATCGCGGTAC (SEQ ID NO:26)
F-I855L	CCGCGATGGCTTCGGCCTCAGCGATAAAG (SEQ ID NO:27)
R-I855L	CTTTATCGCTGAGGCCGAAGCCATCGCGG (SEQ ID NO:28)
F-M593A	GTCAGCGCCCGGTTGAAGCGTACTTTTCTGTTGCAG (SEQ ID NO:29)
R-M593A	CTGCAACAGAAAAGTACGCTTCAACCGGGCGCTGAC (SEQ ID NO:30)
F-M593C	GTCAGCGCCCGGTTGAATGTTACTTTTCTGTTGCAG (SEQ ID NO:31)
R-M593C	CTGCAACAGAAAAGTAACATTCAACCGGGCGCTGAC (SEQ ID NO:32)
F-M593D	GTCAGCGCCCGGTTGAAGACTACTTTTCTGTTGCAG (SEQ ID NO:33)
R-M593D	CTGCAACAGAAAAGTAGTCTTCAACCGGGCGCTGAC (SEQ ID NO:34)
F-M593E	GTCAGCGCCCGGTTGAAGAGTACTTTTCTGTTGCAG (SEQ ID NO:35)
R-M593E	CTGCAACAGAAAAGTACTCTTCAACCGGGCGCTGAC (SEQ ID NO:36)
F-M593F	GTCAGCGCCCGGTTGAATTTTACTTTTCTGTTGCAG (SEQ ID NO:37)
R-M593F	CTGCAACAGAAAAGTAAAATTCAACCGGGCGCTGAC (SEQ ID NO:38)
F-M593G	GTCAGCGCCCGGTTGAAGGGTACTTTTCTGTTGCAG (SEQ ID NO:39)
R-M593G	CTGCAACAGAAAAGTACCCTTCAACCGGGCGCTGAC (SEQ ID NO:40)
F-M593H	FGTCAGCGCCCGGTTGAACACTACTTTTCTGTTGCAG (SEQ ID NO:41)

R-M593H	CTGCAACAGAAAAGTAGTGTTCAACCGGGCGCTGAC (SEQ ID NO:42)
F-M593I	GTCAGCGCCCGGTTGAAATCTACTTTTCTGTTGCAG (SEQ ID NO:43)
R-M593I	CTGCAACAGAAAAGTAGATTTCAACCGGGCGCTGAC (SEQ ID NO:44)
F-M593K	GTCAGCGCCCGGTTGAAAAATACTTTTCTGTTGCAG (SEQ ID NO:45)
R-M593K	CTGCAACAGAAAAGTATTTTTCAACCGGGCGCTGAC (SEQ ID NO:46)
F-M593L	GTCAGCGCCCGGTTGAATTGTACTTTTCTGTTGCAG (SEQ ID NO:47)
R-M593L	CTGCAACAGAAAAGTACAATTCAACCGGGCGCTGAC (SEQ ID NO:48)
F-M593N	GTCAGCGCCCGGTTGAAAACTACTTTTCTGTTGCAG (SEQ ID NO:49)
R-M593N	CTGCAACAGAAAAGTAGTTTTCAACCGGGCGCTGAC (SEQ ID NO:50)
F-M593Q	GTCAGCGCCCGGTTGAACAGTACTTTTCTGTTGCAG (SEQ ID NO:51)
R-M593Q	CTGCAACAGAAAAGTACTGTTCAACCGGGCGCTGAC (SEQ ID NO:52)
F-M593P	GTCAGCGCCCGGTTGAACCGTACTTTTCTGTTGCAG (SEQ ID NO:53)
R-M593P	CTGCAACAGAAAAGTACGGTTCAACCGGGCGCTGAC (SEQ ID NO:54)
F-M593R	GTCAGCGCCCGGTTGAAAGGTACTTTTCTGTTGCAG (SEQ ID NO:55)
R-M593R	CTGCAACAGAAAAGTACCTTTCAACCGGGCGCTGAC (SEQ ID NO:56)
F-M593S	GTCAGCGCCCGGTTGAATCGTACTTTTCTGTTGCAG (SEQ ID NO:57)
R-M593S	CTGCAACAGAAAAGTACGATTCAACCGGGCGCTGAC (SEQ ID NO:58)
F-M593T	GTCAGCGCCCGGTTGAAACGTACTTTTCTGTTGCAG (SEQ ID NO:59)
R-M593T	CTGCAACAGAAAAGTACGTTTCAACCGGGCGCTGAC (SEQ ID NO:60)
F-M593V	GTCAGCGCCCGGTTGAAGTGTACTTTTCTGTTGCAG (SEQ ID NO:61)
R-M593V	CTGCAACAGAAAAGTACACTTCAACCGGGCGCTGAC (SEQ ID NO:62)
F-M593W	GTCAGCGCCCGGTTGAATGGTACTTTTCTGTTGCAG (SEQ ID NO:63)
R-M593W	CTGCAACAGAAAAGTACCATTCAACCGGGCGCTGAC (SEQ ID NO:64)
F-M593Y	GTCAGCGCCCGGTTGAATATTACTTTTCTGTTGCAG (SEQ ID NO:65)
R-M593Y	CTGCAACAGAAAAGTAATATTCAACCGGGCGCTGAC (SEQ ID NO:66)
F-A620C	CCTCTTGCCTGTGCGTAATCCTGGACG (SEQ ID NO:67)
R-A620C	CGTCCAGGATTACGCACAGGCAAGAGG (SEQ ID NO:68)
F-A620D	CCTCTTGCCTGGACGTAATCCTGGACG (SEQ ID NO:69)
R-A620D	CGTCCAGGATTACGTCCAGGCAAGAGG (SEQ ID NO:70)
F-A620E	CCTCTTGCCTGGAAGTAATCCTGGACG (SEQ ID NO:71)
R-A620E	CGTCCAGGATTACTTCCAGGCAAGAGG (SEQ ID NO:72)
F-A620F	CCTCTTGCCTGTTCGTAATCCTGGACG (SEQ ID NO:73)
R-A620F	CGTCCAGGATTACGAACAGGCAAGAGG (SEQ ID NO:74)
F-A620G	CCTCTTGCCTGGGCGTAATCCTGGACG (SEQ ID NO:75)
R-A620G	CGTCCAGGATTACGCCCAGGCAAGAGG (SEQ ID NO:76)
F-A620H	CCTCTTGCCTGCACGTAATCCTGGACG (SEQ ID NO:77)
R-A620H	CGTCCAGGATTACGTGCAGGCAAGAGG (SEQ ID NO:78)
F-A620I	CCTCTTGCCTGATCGTAATCCTGGACG (SEQ ID NO:79)
R-A620I	CGTCCAGGATTACGATCAGGCAAGAGG (SEQ ID NO:80)

F-A620K	CCTCTTGCCTGAAAGTAATCCTGGACG (SEQ ID NO:81)
R-A620K	CGTCCAGGATTACTTTCAGGCAAGAGG (SEQ ID NO:82)
F-A620L	CCTCTTGCCTGCTCGTAATCCTGGACG (SEQ ID NO:83)
R-A620L	CGTCCAGGATTACGAGCAGGCAAGAGG (SEQ ID NO:84)
F- A620M	CCTCTTGCCTGATGGTAATCCTGGACG (SEQ ID NO:85)
R-A620M	CGTCCAGGATTACCATCAGGCAAGAGG (SEQ ID NO:86)
F-A620N	CCTCTTGCCTGAACGTAATCCTGGACG (SEQ ID NO:87)
R-A620N	CGTCCAGGATTACGTTCAGGCAAGAGG (SEQ ID NO:88)
F-A620P	CCTCTTGCCTGCCCGTAATCCTGGACG (SEQ ID NO:89)
R-A620P	CGTCCAGGATTACGGGCAGGCAAGAGG (SEQ ID NO:90)
F-A620Q	CCTCTTGCCTGCAAGTAATCCTGGACG (SEQ ID NO:91)
R-A620Q	CGTCCAGGATTACTTGCAGGCAAGAGG (SEQ ID NO:92)
F-A620R	CCTCTTGCCTGCGCGTAATCCTGGACG (SEQ ID NO:93)
R-A620R	CGTCCAGGATTACGCGCAGGCAAGAGG (SEQ ID NO:94)
F-A620S	CCTCTTGCCTGTCCGTAATCCTGGACG (SEQ ID NO:95)
R-A620S	CGTCCAGGATTACGGACAGGCAAGAGG (SEQ ID NO:96)
F-A620T	CCTCTTGCCTGACCGTAATCCTGGACG (SEQ ID NO:97)
R-A620T	CGTCCAGGATTACGGTCAGGCAAGAGG (SEQ ID NO:98)
F-A620V	CCTCTTGCCTGGTCGTAATCCTGGACG (SEQ ID NO:99)
R-A620V	CGTCCAGGATTACGACCAGGCAAGAGG (SEQ ID NO:100)
F-A620W	CCTCTTGCCTGTGGGTAATCCTGGACG (SEQ ID NO:101)
R-A620W	CGTCCAGGATTACCCACAGGCAAGAGG (SEQ ID NO:102)
F-A620Y	CCTCTTGCCTGTACGTAATCCTGGACG (SEQ ID NO:103)
R-A620Y	CGTCCAGGATTACGTACAGGCAAGAGG (SEQ ID NO:104)
F-Y700A	GGCCTGCTGGCCTCCGCTACCAAGGAAGCG (SEQ ID NO:105)
R-Y700A	CGCTTCCTTGGTAGCGGAGGCCAGCAGGCC (SEQ ID NO:106)
F-Y700C	GGCCTGCTGGCCTCCTGTACCAAGGAAGCG (SEQ ID NO:107)
R-Y700C	CGCTTCCTTGGTACAGGAGGCCAGCAGGCC (SEQ ID NO:108)
F-Y700D	GGCCTGCTGGCCTCCGATACCAAGGAAGCG (SEQ ID NO:109)
R-Y700D	CGCTTCCTTGGTATCGGAGGCCAGCAGGCC (SEQ ID NO:110)
F-Y700E	GGCCTGCTGGCCTCCGAAACCAAGGAAGCG (SEQ ID NO:111)
R-Y700E	CGCTTCCTTGGTTTCGGAGGCCAGCAGGCC (SEQ ID NO:112)
F-Y700F	GGCCTGCTGGCCTCCTTTACCAAGGAAGCG (SEQ ID NO:113)
R-Y700F	CGCTTCCTTGGTAAAGGAGGCCAGCAGGCC (SEQ ID NO:114)
F-Y700G	GGCCTGCTGGCCTCCGGTACCAAGGAAGCG (SEQ ID NO:115)
R-Y700G	CGCTTCCTTGGTACCGGAGGCCAGCAGGCC (SEQ ID NO:116)
F-Y700H	GGCCTGCTGGCCTCCCATACCAAGGAAGCG (SEQ ID NO:117)
R-Y700H	CGCTTCCTTGGTATGGGAGGCCAGCAGGCC (SEQ ID NO:118)
F-Y700I	GGCCTGCTGGCCTCCATTACCAAGGAAGCG (SEQ ID NO:119)

R-Y700I	CGCTTCCTTGGTAATGGAGGCCAGCAGGCC (SEQ ID NO:120)
F-Y700K	GGCCTGCTGGCCTCCAAAACCAAGGAAGCG (SEQ ID NO:121)
R-Y700K	CGCTTCCTTGGTTTTGGAGGCCAGCAGGCC (SEQ ID NO:122)
F-Y700L	GGCCTGCTGGCCTCCTTAACCAAGGAAGCG (SEQ ID NO:123)
R-Y700L	CGCTTCCTTGGTTAAGGAGGCCAGCAGGCC (SEQ ID NO:124)
F-Y700M	GGCCTGCTGGCCTCCATGACCAAGGAAGCG (SEQ ID NO:125)
R-Y700M	CGCTTCCTTGGTCATGGAGGCCAGCAGGCC (SEQ ID NO:126)
F-Y700N	GGCCTGCTGGCCTCCAATACCAAGGAAGCG (SEQ ID NO:127)
R-Y700N	CGCTTCCTTGGTATTGGAGGCCAGCAGGCC (SEQ ID NO:128)
F-Y700P	GGCCTGCTGGCCTCCCCTACCAAGGAAGCG (SEQ ID NO:129)
R-Y700P	CGCTTCCTTGGTAGGGGAGGCCAGCAGGCC (SEQ ID NO:130)
F-Y700Q	GGCCTGCTGGCCTCCGAAACCAAGGAAGCG (SEQ ID NO:131)
R-Y700Q	CGCTTCCTTGGTTTCGGAGGCCAGCAGGCC (SEQ ID NO:132)
F-Y700R	GGCCTGCTGGCCTCCCGTACCAAGGAAGCG (SEQ ID NO:133)
RY700R	CGCTTCCTTGGTACGGGAGGCCAGCAGGCC (SEQ ID NO:134)
F-Y700S	GGCCTGCTGGCCTCCTCTACCAAGGAAGCG (SEQ ID NO:135)
R-Y700S	CGCTTCCTTGGTAGAGGAGGCCAGCAGGCC (SEQ ID NO:136)
F-Y700T	GGCCTGCTGGCCTCCACTACCAAGGAAGCG (SEQ ID NO:137)
R-Y700T	CGCTTCCTTGGTAGTGGAGGCCAGCAGGCC (SEQ ID NO:138)
F-Y700V	GGCCTGCTGGCCTCCGTTACCAAGGAAGCG (SEQ ID NO:139)
R-Y700V	CGCTTCCTTGGTAACGGAGGCCAGCAGGCC (SEQ ID NO:140)
F-Y700W	GGCCTGCTGGCCTCCTGGACCAAGGAAGCG (SEQ ID NO:141)
R-Y700W	CGCTTCCTTGGTCCAGGAGGCCAGCAGGCC (SEQ ID NO:142)

The letter F and R in the beginning of each mutagenic oligonucleotide indicates 'forward' and 'reverse' sequence, respectively.

References

1. van Beek, T. A. & Montoro, P. Chemical analysis and quality control of Ginkgo biloba leaves, extracts, and phytopharmaceuticals. J Chromatogr A 1216, 2002-32 (2009).

2. Aponte, M. et al. Activation of platelet-activating factor receptor and pleiotropic effects on tyrosine phospho-EGFR/Src/FAK/paxillin in ovarian cancer. Cancer Res 68, 5839-48 (2008).

3. Heads, J. A., Hawthorne, R. L., Lynagh, T. & Lynch, J. W. Structure-activity analysis of ginkgolide binding in the glycine receptor pore. J Neurochem 105, 1418-27 (2008).

4. Ivic, L. et al. Terpene trilactones from Ginkgo biloba are antagonists of cortical glycine and GABA(A) receptors. J Biol Chem 278, 49279-85 (2003).

5. Jensen, A. A. et al. Probing the pharmacophore of ginkgolides as glycine receptor antagonists. J Med Chem 50, 1610-7 (2007).

6. Ye, B. et al. Ginkgo biloba and ovarian cancer prevention: epidemiological and biological evidence. Cancer Lett 251, 43-52 (2007).

7. Kang, S. M. et al. Effect of supplementing terpenoid biosynthetic precursors on the accumulation of bilobalide and ginkgolides in Ginkgo biloba cell cultures. J Biotechnol 123, 85-92 (2006).

8. Crimmins, M. T. et al. The total synthesis of (+/-)-ginkgolide B. Journal of the American Chemical Society 122, 8453-8463 (2000).

9. Carter, O. A., Peters, R. J. & Croteau, R. Monoterpene biosynthesis pathway construction in Escherichia coli. Phytochemistry 64, 425-33 (2003).

10. Engels, B., Dahm, P. & Jennewein, S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. Metab Eng 10, 201-6 (2008).

11. Huang, Q., Roessner, C. A., Croteau, R. & Scott, A. I. Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. Bioorg Med Chem 9, 2237-42 (2001).

12. Reiling, K. K. et al. Mono and diterpene production in Escherichia coli. Biotechnol Bioeng 87, 200-12 (2004).

13. Ro, D. K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-3 (2006).

14. Leonard, E., Runguphan, W., O'Connor, S. & Prather, K. J. Opportunities in metabolic engineering to facilitate scalable alkaloid production. Nat Chem Biol 5, 292-300 (2009).

15. Roberts, S. C. Production and engineering of terpenoids in plant cell culture. Nat Chem Biol 3, 387-95 (2007).

16. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23, 612-6 (2005).

17. Conrado, R. J., Varner, J. D. & DeLisa, M. P. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol 19, 492-9 (2008).

18. Fong, S. S., Nanchen, A., Palsson, B. O. & Sauer, U. Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes. J Biol Chem 281, 8024-33 (2006).

19. Copley, S. D. Evolution of efficient pathways for degradation of anthropogenic chemicals. Nat Chem Biol 5, 559-66 (2009).

20. Fischbach, M. A. & Clardy, J. One pathway, many products. Nat Chem Biol 3, 353-5 (2007).

21. Jensen, R. A. Enzyme recruitment in evolution of new function. Annu Rev Microbiol 30, 409-25 (1976).

22. Le Rouzic, A. & Carlborg, O. Evolutionary potential of hidden genetic variation. Trends Ecol Evol 23, 33-7 (2008).

23. Sniegowski, P. D. & Murphy, H. A. Evolvability. Curr Biol 16, R831-4 (2006).

24. Aharoni, A. et al. The 'evolvability' of promiscuous protein functions. Nat Genet 37, 73-6 (2005).

25. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol 13, 3-9 (2009).

26. O'Maille, P. E. et al. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. Nat Chem Biol 4, 617-23 (2008).

27. Ohnuma, S. et al. A pathway where polyprenyl diphosphate elongates in prenyltransferase. Insight into a common mechanism of chain length determination of prenyltransferases. J Biol Chem 273, 26705-13 (1998).

28. Peters, R. J. et al. Abietadiene synthase from grand fir (Abies grandis): characterization and mechanism of action of the "pseudomature" recombinant enzyme. Biochemistry 39, 15592-602 (2000).

29. Ravn, M. M., Coates, R. M., Flory, J. E., Peters, R. J. & Croteau, R. Stereochemistry of the cyclization-rearrangement of (+)-copalyl diphosphate to (-)-abietadiene catalyzed by recombinant abietadiene synthase from Abies grandis. Org Lett 2, 573-6 (2000).

30. Starks, C. M., Back, K., Chappell, J. & Noel, J. P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277, 1815-20 (1997).

31. Keeling, C. I., Weisshaar, S., Lin, R. P. & Bohlmann, J. Functional plasticity of paralogous diterpene synthases involved in conifer defense. Proc Natl Acad Sci U S A 105, 1085-90 (2008).

32. Peters, R. J. & Croteau, R. B. Abietadiene synthase catalysis: mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. Proc Natl Acad Sci U S A 99, 580-4 (2002).

33. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. Nature 440, 1078-82 (2006).

34. Greenhagen, B. T., O'Maille, P. E., Noel, J. P. & Chappell, J. Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proc Natl Acad Sci U S A 103, 9826-31 (2006).

35. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157, 105-32 (1982).

36. Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. & Hodges, R. S. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. J Pept Sci 1, 319-29 (1995).

37. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. Proc Natl Acad Sci U S A 106 Suppl 1, 9995-10000 (2009).

38. Li, Y. et al. A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. Nat Biotechnol 25, 1051-6 (2007).

39. Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. Biol Direct 2, 17 (2007).

40. Nims, E., Dubois, C. P., Roberts, S. C. & Walker, E. L. Expression profiling of genes involved in paclitaxel biosynthesis for targeted metabolic engineering. Metab Eng 8, 385-94 (2006).

41. Kloer, D. P., Welsch, R., Beyer, P. & Schulz, G. E. Structure and reaction geometry of geranylgeranyl diphosphate synthase from Sinapis alba. Biochemistry 45, 15197-204 (2006).

42. Liao, Z. et al. A new geranylgeranyl diphosphate synthase gene from Ginkgo biloba, which intermediates the biosynthesis of the key precursor for ginkgolides. DNA Seq 15, 153-8 (2004).

43. Hefner, J., Ketchum, R. E. B. & Croteau, R. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from Taxus canadensis and assessment of the role of this prenyltransferase in cells induced for Taxol production. Archives of Biochemistry and Biophysics 360, 62-74 (1998).

44. Hosfield, D. J. et al. Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. J Biol Chem 279, 8526-9 (2004).

45. Trapp, S. C. & Croteau, R. B. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 158, 811-32 (2001).

46. Gonzalez, M. A., Correa-Royero, J., Agudelo, L., Mesa, A. & Betancur-Galvis, L. Synthesis and biological evaluation of abietic acid derivatives. Eur J Med Chem 44, 2468-72 (2009).

47. Li, J. W. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? Science 325, 161-5 (2009).

48. Dueber, J. E. et al. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27, 753-9 (2009).

49. Schepmann, H. G., Pang, J. & Matsuda, S. P. Cloning and characterization of Ginkgo biloba levopimaradiene synthase which catalyzes the first committed step in ginkgolide biosynthesis. Arch Biochem Biophys 392, 263-9 (2001).

50. Leonard, E. & Koffas, M. A. Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by Escherichia coli. Appl Environ Microbiol 73, 7246-51 (2007).

51. Cyr, A., Wilderman, P. R., Determan, M. & Peters, R. J. A modular approach for facile biosynthesis of labdane-related diterpenes. J Am Chem Soc 129, 6684-5 (2007).

52. Martin, D. M., Faldt, J. & Bohlmann, J. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. Plant Physiol 135, 1908-27 (2004).

53. Facchini, P. J. & Chappell, J. Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. Proc Natl Acad Sci U S A 89, 11088-92 (1992).

54. Hill, A. M., Cane, D. E., Mau, C. J. & West, C. A. High level expression of Ricinus communis casbene synthase in Escherichia coli and characterization of the recombinant enzyme. Arch Biochem Biophys 336, 283-9 (1996).

55. Huang, K. X., Huang, Q. L., Wildung, M. R., Croteau, R. & Scott, A. I. Overproduction, in Escherichia coli, of soluble taxadiene synthase, a key enzyme in the Taxol biosynthetic pathway. Protein Expr Purif 13, 90-6 (1998).

56. Morrone, D. et al. An unexpected diterpene cyclase from rice: functional identification of a stemodene synthase. Arch Biochem Biophys 448, 133-40 (2006).

57. Toyomasu, T. Recent advances regarding diterpene cyclase genes in higher plants and fungi. Biosci Biotechnol Biochem 72, 1168-75 (2008).

58. Dairi, T. et al. Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic terpentecin. J Bacteriol 183, 6085-94 (2001).

Having thus described several aspects of at least one embodiment of this invention, it is to be appreciated various alterations, modifications, and improvements will readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this disclosure, and are intended to be within the spirit and scope of the invention. Accordingly, the foregoing description and drawings are by way of example only. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All references disclosed herein are incorporated by reference in their entirety for the specific purpose mentioned herein.

What is claimed is:

Claims

1. A method comprising:

recombinantly expressing a terpenoid synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme in a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway.

2. The method of claim 1, wherein the cell is a bacterial cell.

3. The method of claim 2, wherein the cell is an *Escherichia coli* cell.

4. The method of claim 2, wherein the cell is a Gram-positive cell.

5. The method of claim 4, wherein the cell is a *Bacillus* cell.

6. The method of claim 1, wherein the cell is a yeast cell.

7. The method of claim 6, wherein the yeast cell is a *Saccharomyces* cell.

8. The method of claim 6, wherein the yeast cell is a *Yarrowia* cell.

9. The method of claim 1, wherein the cell is an algal cell.

10. The method of claim 1, wherein the cell is a plant cell.

11. The method of any one of claims 1-10, wherein the terpenoid synthase enzyme is a diterpenoid synthase enzyme.

12. The method of claim 11, wherein the diterpenoid synthase enzyme is a levopimaradiene synthase (LPS) enzyme.

13. The method of claim 12, wherein the LPS enzyme is a *Ginkgo biloba* enzyme.

14. The method of claim 12 or 13, wherein the LPS enzyme contains one or more mutations.

15. The method of claim 14, wherein the LPS enzyme contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme.

16. The method of claim 15, wherein the LPS enzyme contains one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

17. The method of claim 16, wherein the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

18. The method of any one of claims 1-17, wherein the GGPPS enzyme is a *Taxus canadensis* enzyme.

19. The method of any one of claims 1-18, wherein the GGPPS enzyme contains one or more mutations.

20. The method of claim 19, wherein the GGPPS enzyme contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme.

21. The method of claim 20, wherein the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

22. The method of any one of claims 12-21, wherein the LPS enzyme contains the mutation M593I and/or Y700F, corresponding to residues within the full-length wild-type *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme, and the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or more equivalent mutations in a homologous GGPPS enzyme.

23. The method of any one of claims 1-22, wherein the gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme is expressed from one or more plasmids.

24. The method of any one of claims 1-23, wherein the gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme is incorporated into the genome of the cell.

25. The method of any one of claims 1-24, further comprising culturing the cell.

26. The method of any one of claims 1-25, wherein the cell produces a terpenoid.

27. The method of claim 26, wherein the terpenoid has one or more cyclic structures.

28. The method of claim 26 or 27, wherein the terpenoid is a diterpenoid.

29. The method of claim 28, wherein the diterpenoid is levopimaradiene.

30. The method of any one of claims 25-29, further comprising recovering the terpenoid from the cell culture.

31. The method of claim 30, wherein the terpenoid is recovered from the gas phase.

32. The method of claim 30, wherein an organic layer is added to the cell culture, and the terpenoid is recovered from the organic layer.

33. The method of any one of claims 1-32, wherein the cell produces a Taxol, a gibberellin, and/or a steviol glycoside.

34. The method of any one of claims 1-33, wherein the terpenoid synthase enzyme and/or the geranylgeranyl diphosphate synthase (GGPPS) enzyme is codon optimized.

35. A cell that overexpresses one or more components of the non-mevalonate (MEP) pathway, and that recombinantly expresses a terpenoid synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme.

36. The cell of claim 35, wherein the cell is a bacterial cell.

37. The cell of claim 36, wherein the cell is an *Escherichia coli* cell.

38. The cell of claim 36, wherein the cell is a Gram-positive cell.

39. The cell of claim 38, wherein the cell is a *Bacillus* cell.

40. The cell of claim 35, wherein the cell is a yeast cell.

41. The cell of claim 40, wherein the yeast cell is a *Saccharomyces* cell.

42. The cell of claim 40, wherein the yeast cell is a Yarrowia cell.

43. The cell of claim 35, wherein the cell is an algal cell.

44. The cell of claim 35, wherein the cell is a plant cell.

45. The cell of any one of claims 35-44, wherein the terpenoid synthase enzyme is a diterpenoid synthase enzyme.

46. The cell of claim 45, wherein the diterpenoid synthase enzyme is a levopimaradiene synthase (LPS) enzyme.

47. The cell of claim 46, wherein the LPS enzyme is a *Ginkgo biloba* enzyme.

48. The cell of claim 46 or 47, wherein the LPS enzyme contains one or more mutations.

49. The cell of claim 48, wherein the LPS enzyme contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme.

50. The cell of claim 49, wherein the LPS enzyme contains one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

51. The cell of claim 50, wherein the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

52. The cell of any one of claims 35-51, wherein the GGPPS enzyme is a *Taxus canadensis* enzyme.

53. The cell of any one of claims 35-52, wherein the GGPPS enzyme contains one or more mutations.

54. The cell of claim 53, wherein the GGPPS enzyme contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme.

55. The cell of claim 54, wherein the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

56. The cell of any one of claims 46-55, wherein the LPS enzyme contains the mutation M593I and/or Y700F, corresponding to residues within the full-length wild-type *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme, and the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or more equivalent mutations in a homologous GGPPS enzyme.

57. The cell of any one of claims 35-56, wherein the gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme is expressed from one or more plasmids.

58. The cell of any one of claims 35-57, wherein the gene encoding for the terpenoid synthase enzyme and/or the gene encoding for the geranylgeranyl diphosphate synthase (GGPPS) enzyme is incorporated into the genome of the cell.

59. The cell of any one of claims 35-58, wherein the cell produces a terpenoid.

60. The cell of claim 59, wherein the terpenoid has one or more cyclic structures.

61. The cell of claim 59 or 60, wherein the terpenoid is a diterpenoid.

62. The cell of claim 61, wherein the diterpenoid is levopimaradiene.

63. The cell of any one of claims 35-62, wherein the cell produces a Taxol, a gibberellin, and/or a steviol glycoside.

64. The cell of any one of claims 35-63, wherein the terpenoid synthase enzyme and/or the geranylgeranyl diphosphate synthase (GGPPS) enzyme is codon-optimized.

65. A cell that recombinantly expresses a levopimaradiene synthase (LPS) enzyme, wherein the LPS enzyme contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more mutations in equivalent residues within a homologous LPS enzyme.

66. The cell of claim 65, wherein the LPS enzyme contains one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

67. The cell of claim 66, wherein the LPS enzyme contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F, corresponding to mutations at residues within the full-length, wild-type, *Ginkgo biloba* LPS enzyme, or one or more equivalent mutations in a homologous LPS enzyme.

68. The cell of any one of claims 65-67, wherein the cell is a bacterial cell.

69. The cell of claim 68, wherein the cell is an *Escherichia coli* cell.

70. The cell of claim 68, wherein the cell is a Gram-positive cell.

71. The cell of claim 70, wherein the cell is a *Bacillus* cell.

72. The cell of any one of claims 65-67, wherein the cell is a yeast cell.

73. The cell of claim 72, wherein the yeast cell is a *Saccharomyces* cell.

74. The cell of claim 72, wherein the yeast cell is a Yarrowia cell.

75. The cell of any one of claims 65-67, wherein the cell is an algal cell.

76. The cell of any one of claims 65-67, wherein the cell is a plant cell.

77. The cell of any one of claims 65-76, wherein the LPS enzyme is codon optimized.

78. A cell that recombinantly expresses a geranylgeranyl diphosphate synthase (GGPPS) enzyme, wherein the GGPPS enzyme contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or a mutation in one or both equivalent residues within a homologous GGPPS enzyme.

79. The cell of claim 78, wherein the GGPPS enzyme contains the mutation S239C and/or G295D, corresponding to mutations at residues within the full-length, wild-type, *Taxus canadensis* GGPPS enzyme, or one or both equivalent mutations in a homologous GGPPS enzyme.

80. The cell of claim 78 or 79, wherein the cell is a bacterial cell.

81. The cell of claim 80, wherein the cell is an *Escherichia coli* cell.

82. The cell of claim 80, wherein the cell is a Gram-positive cell.

83. The cell of claim 82, wherein the cell is a *Bacillus* cell.

84. The cell of claim 78 or 79, wherein the cell is a yeast cell.

85. The cell of claim 84, wherein the yeast cell is a *Saccharomyces* cell.

86. The cell of claim 84, wherein the yeast cell is a Yarrowia cell.

87. The cell of claim 78 or 79, wherein the cell is an algal cell.

88. The cell of claim 78 or 79, wherein the cell is a plant cell.

89. The cell of any one of claims 78-88, wherein the GGPPS enzyme is codon-optimized.

90. An isolated levopimaradiene synthase (LPS) polypeptide that contains a mutation at one or more of the residues selected from the group consisting of: M593, C618, A620, L696, Y700, K723, A729, V731, N838, and I855, corresponding to residues within the full-length, wild-type, *Ginkgo biloba* LPS polypeptide (GenBank Accession No. AF331704).

91. The isolated LPS polypeptide of claim 90, wherein the LPS polypeptide contains one or more mutations selected from the group consisting of: M593I, M593L, C618N, L696Q, K723S, V731L, N838E, I855L, A729G, Y700H, Y700A, Y700C, Y700F, Y700M, Y700W, A620C, A620G, A620S, A620T and A620V.

92. The isolated LPS polypeptide of claim 91, wherein the LPS polypeptide contains the mutation M593I and one of the mutations selected from the group consisting of Y700A, Y700C and Y700F.

93. The isolated LPS polypeptide of any one of claims 90-92, wherein the isolated LPS polypeptide is codon-optomized.

94. An isolated nucleic acid molecule encoding the LPS polypeptide of any one of claims 90-93.

95. A recombinant expression vector comprising the nucleic acid molecule of claim 94 operably linked to a transcription regulatory element.

96. A library comprising the nucleic acid molecule of claim 94.

97. A library comprising the LPS polypeptide of any one of claims 90-93.

98. An isolated geranylgeranyl diphosphate synthase (GGPPS) polypeptide, wherein the GGPPS polypeptide contains a mutation at residue S239 and/or G295, corresponding to residues within the full-length, wild-type, *Taxus canadensis* GGPPS polypeptide (GenBank Accession No. AF081514).

99. The isolated GGPPS polypeptide of claim 98, wherein the GGPPS polypeptide contains the mutation S239C and/or the mutation G295D.

100. The isolated GGPPS polypeptide of claim 98 or 99, wherein the isolated GGPPS polypeptide is codon-optimized.

101. An isolated nucleic acid molecule encoding the GGPPS polypeptide of any one of claims 98-100.

102. A recombinant expression vector comprising the nucleic acid molecule of claim 100, operably linked to a transcription regulatory element.

103. A library comprising the nucleic acid molecule of claim 100.

104. A library comprising the GGPPS polypeptide of any one of claims 98-100.

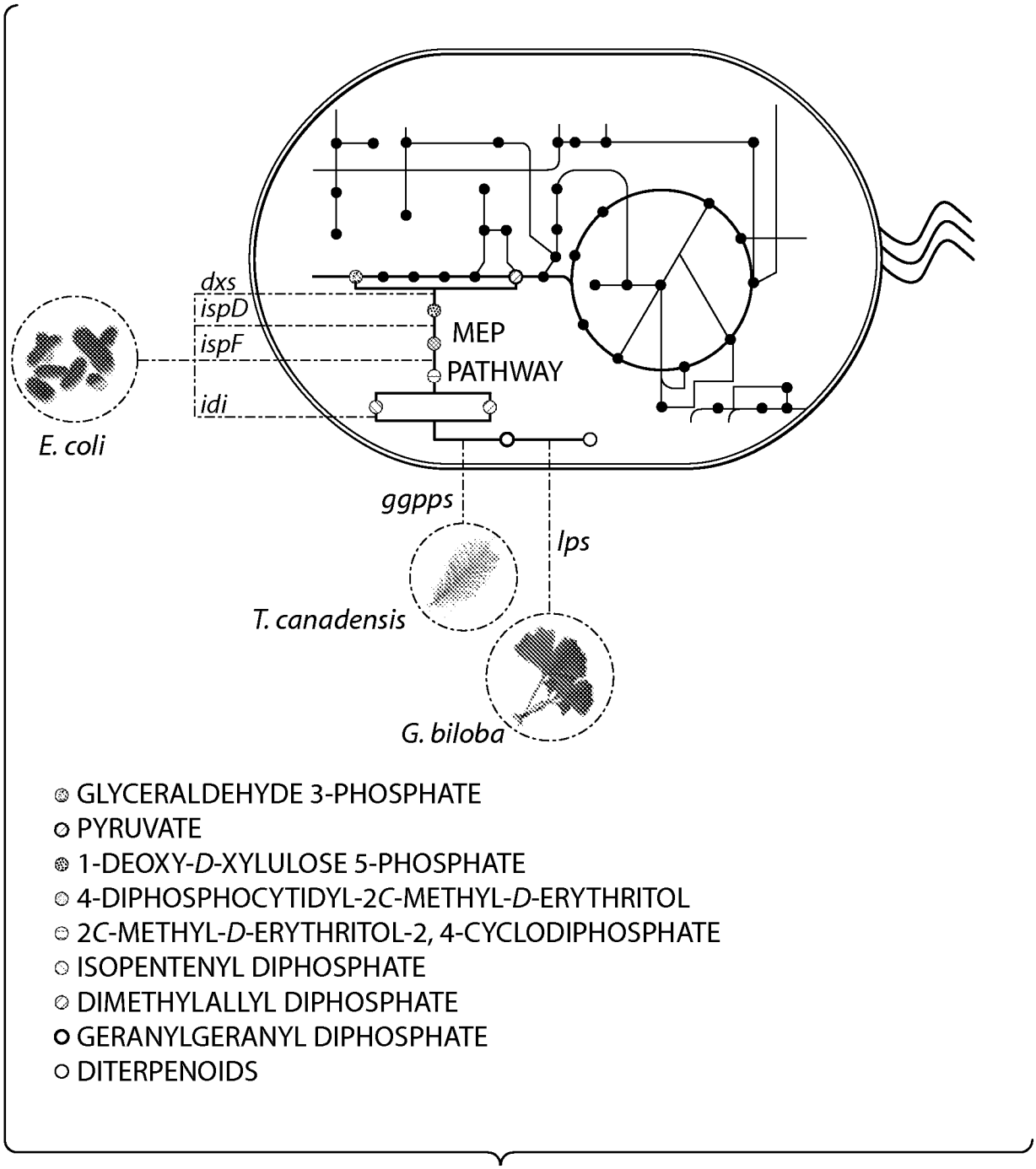

dxs
ispD
ispF
idi
MEP
PATHWAY
E. coli
ggpps
lps
T. canadensis
G. biloba
GLYCERALDEHYDE 3-PHOSPHATE
PYRUVATE
1-DEOXY-D-XYLULOSE 5-PHOSPHATE
4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL
2C-METHYL-D-ERYTHRITOL-2, 4-CYCLODIPHOSPHATE
ISOPENTENYL DIPHOSPHATE
DIMETHYLALLYL DIPHOSPHATE
GERANYLGERANYL DIPHOSPHATE
DITERPENOIDS

Fig. 1A

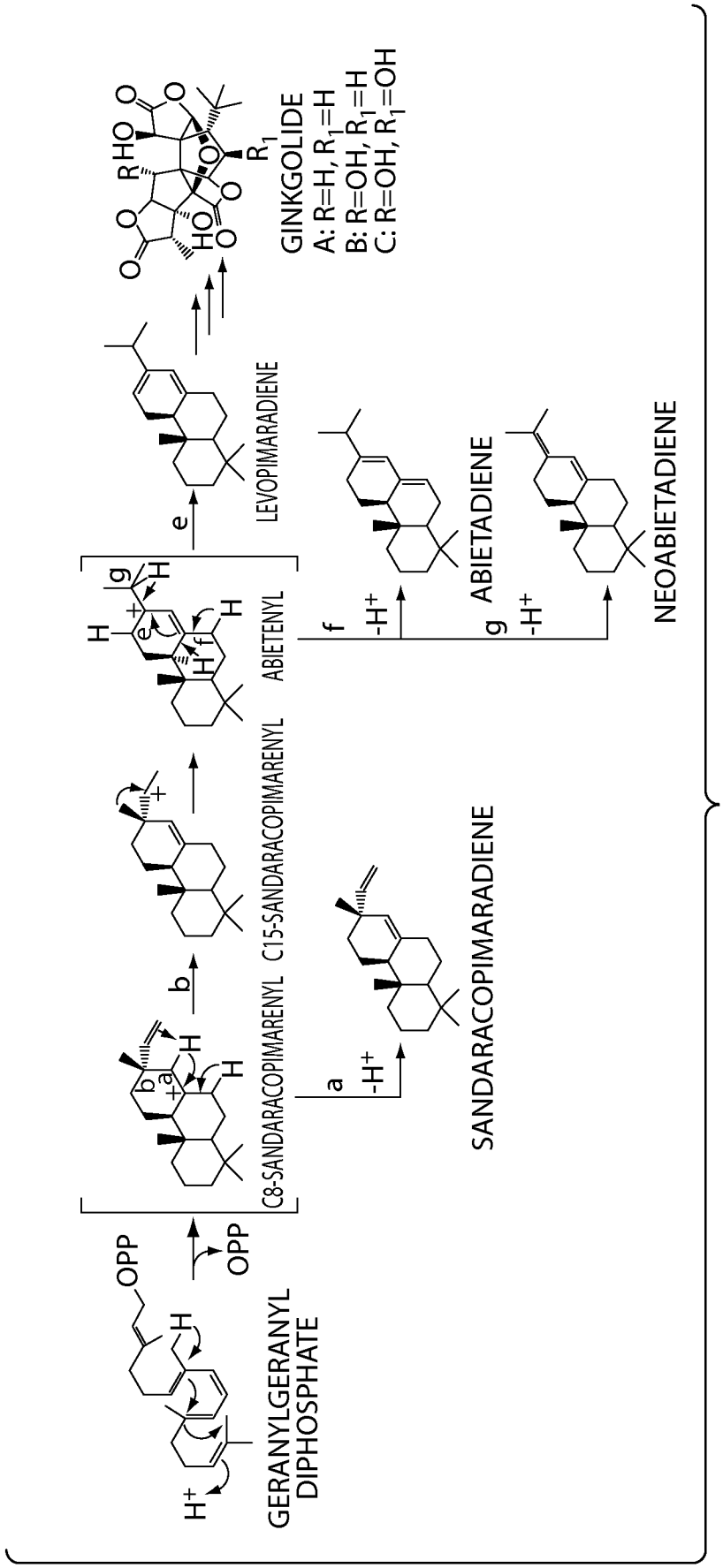

GERANYLGERANYL DIPHOSPHATE
OPP
OPP
H+
C8-SANDARACOPIMARENYL
C15-SANDARACOPIMARENYL
ABIETENYL
a
b
e
f
g
-H+
SANDARACOPIMARADIENE
ABIETADIENE
NEOABIETADIENE
LEVOPIMARADIENE
GINKGOLIDE
A: R=H, R1=H
B: R=OH, R1=H
C: R=OH, R1=OH

Fig. 1B

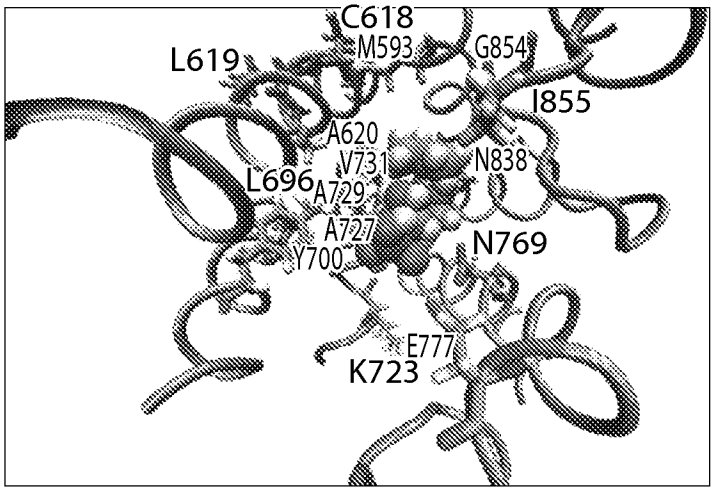


Fig. 2A

POSITION	GbLPS	AbAS	PaAS	PaISO
593	M	I	I	I
618	C	N	N	I
619	L	F	F	F
620	A	T	T	T
696	L	Q	Q	L
700	Y	Y	Y	H
723	K	S	S	S
727	A	A	A	S
729	A	G	G	G
731	V	V	V	L
769	N	N	N	N
777	E	E	E	E
838	N	E	E	E
854	G	T	T	T
855	I	L	L	L

Fig. 2B

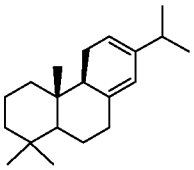
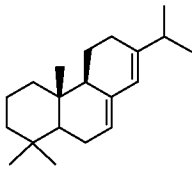
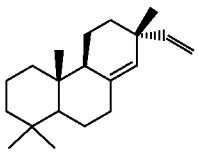
MUTATION	 1	 2	 3	PRODUCTIVITY (FOLD)
M593I	83	12	5	3.7
C618N	91	9	TA	0.2
L619F	94	6	TA	1.5
A620T	87	11	2	1.3
L696Q	86	7	7	1.6
Y700H	71	TA	29	0.4
K723S	90	7	3	1.8
A727S	ND	ND	ND	0.0
A729G	ND	ND	100	0.02
V731L	100	ND	ND	0.08
N769A	ND	ND	ND	0.0
E777A	ND	ND	ND	0.0
N838E	92	6	2	2.2
G854T	79	14	7	1.4
I855L	90	7	3	0.7

Fig. 2C

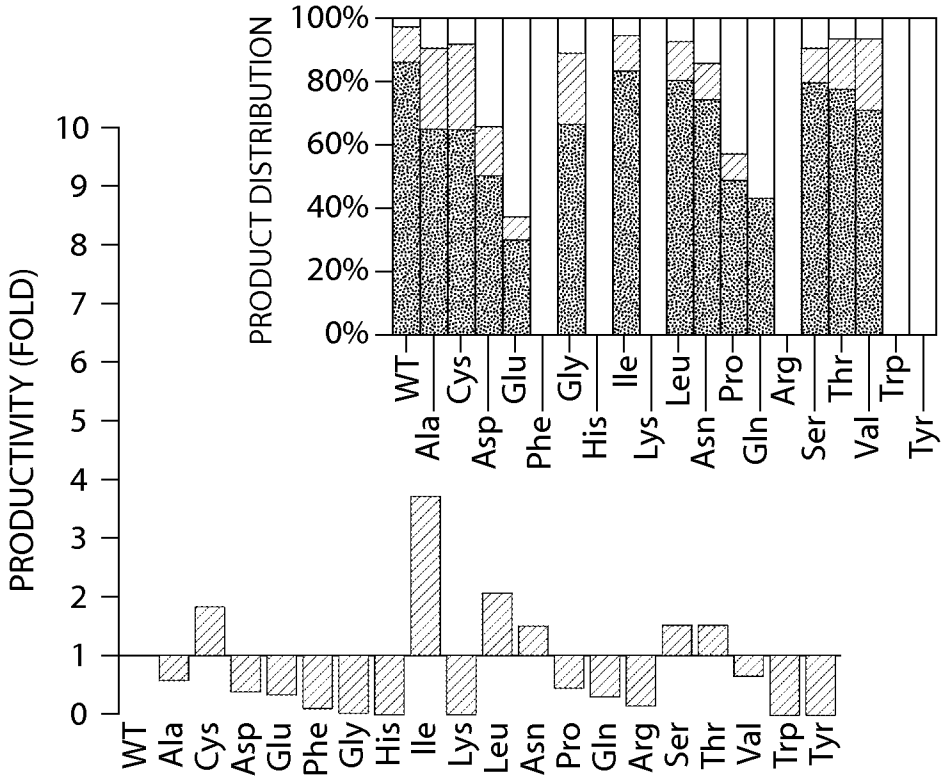

PRODUCTIVITY (FOLD)
PRODUCT DISTRIBUTION
100%
80%
60%
40%
20%
0%
WT Ala Cys Asp Glu Phe Gly His Ile Lys Leu Asn Pro Gln Arg Ser Thr Val Trp Tyr

Fig. 3A

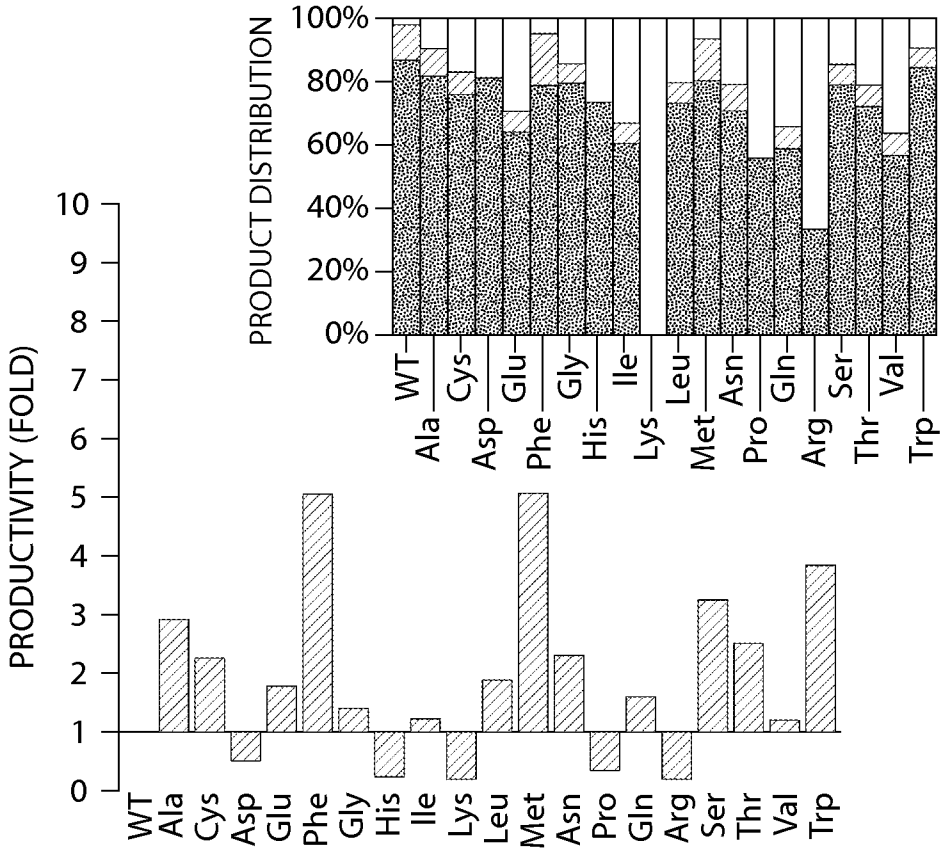

PRODUCTIVITY (FOLD)
PRODUCT DISTRIBUTION
100%
80%
60%
40%
20%
0%
WT Ala Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp

Fig. 3B

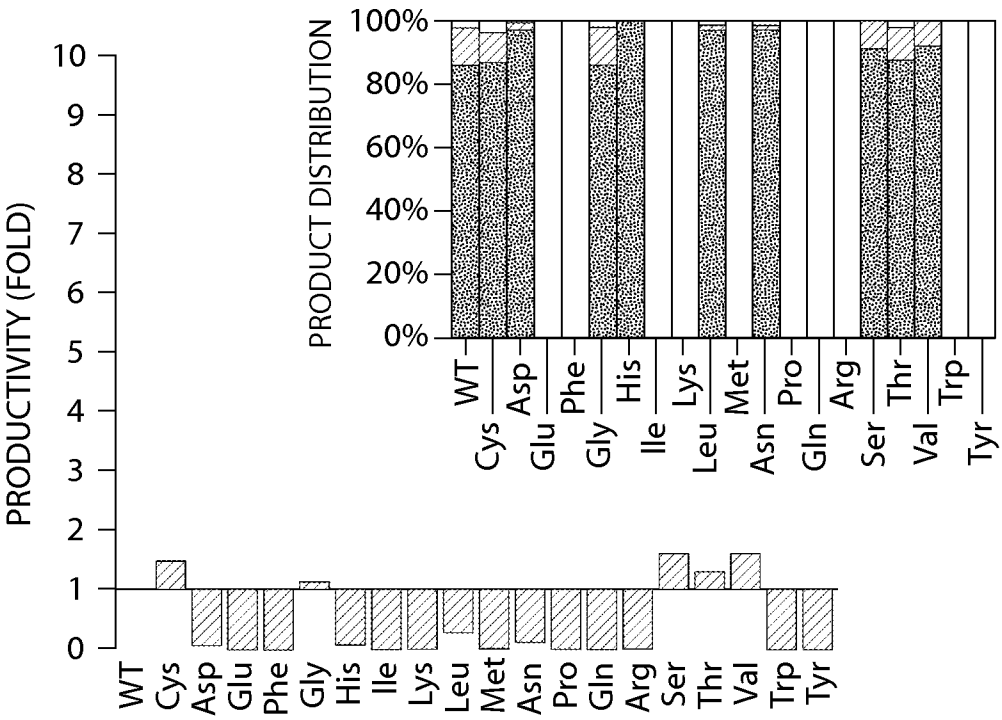

PRODUCTIVITY (FOLD)
PRODUCT DISTRIBUTION
100%
80%
60%
40%
20%
0%
WT Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp Tyr

Fig. 3C

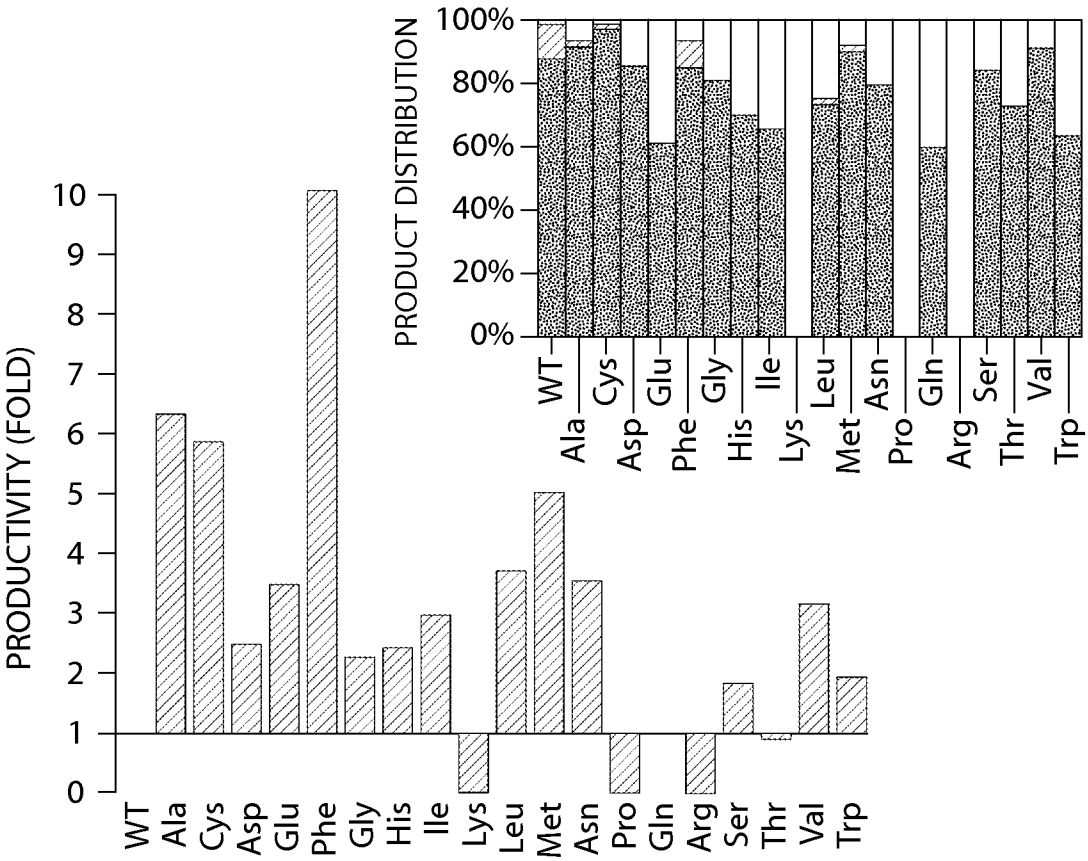

PRODUCTIVITY (FOLD)
PRODUCT DISTRIBUTION
100%
80%
60%
40%
20%
0%
WT Ala Cys Asp Glu Phe Gly His Ile Lys Leu Met Asn Pro Gln Arg Ser Thr Val Trp

Fig. 3D

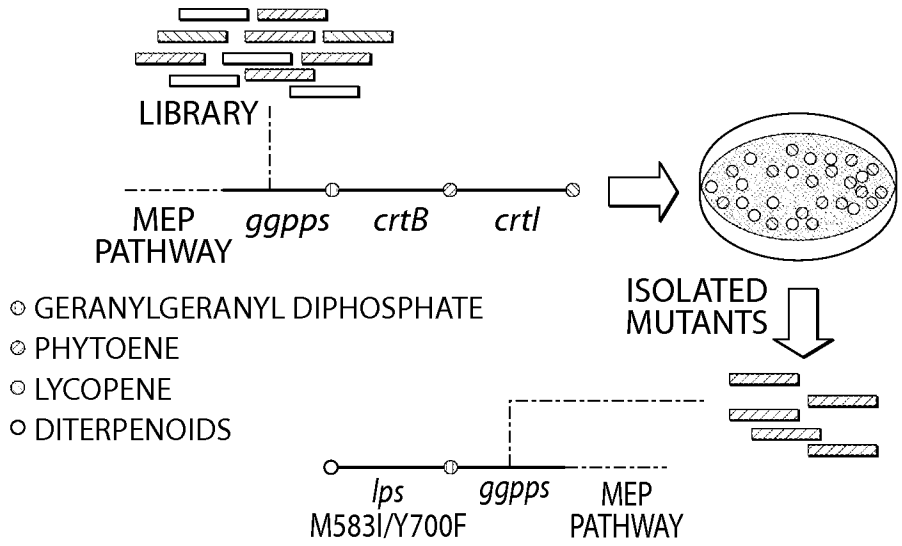

LIBRARY
MEP PATHWAY
ggpps
crtB
crtI
ISOLATED MUTANTS
GERANYLGERANYL DIPHOSPHATE
PHYTOENE
LYCOPENE
DITERPENOIDS
lps M583I/Y700F
ggpps
MEP PATHWAY

Fig. 4A

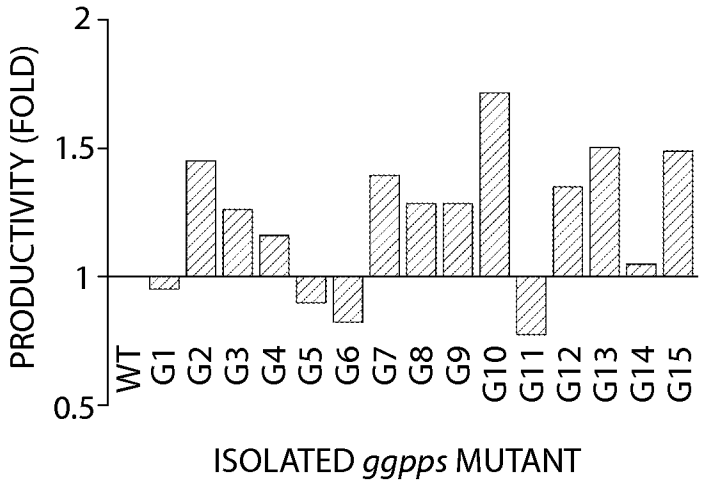

PRODUCTIVITY (FOLD)
2
1.5
1
0.5
WT
G1
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12
G13
G14
G15
ISOLATED ggpps MUTANT

Fig. 4B

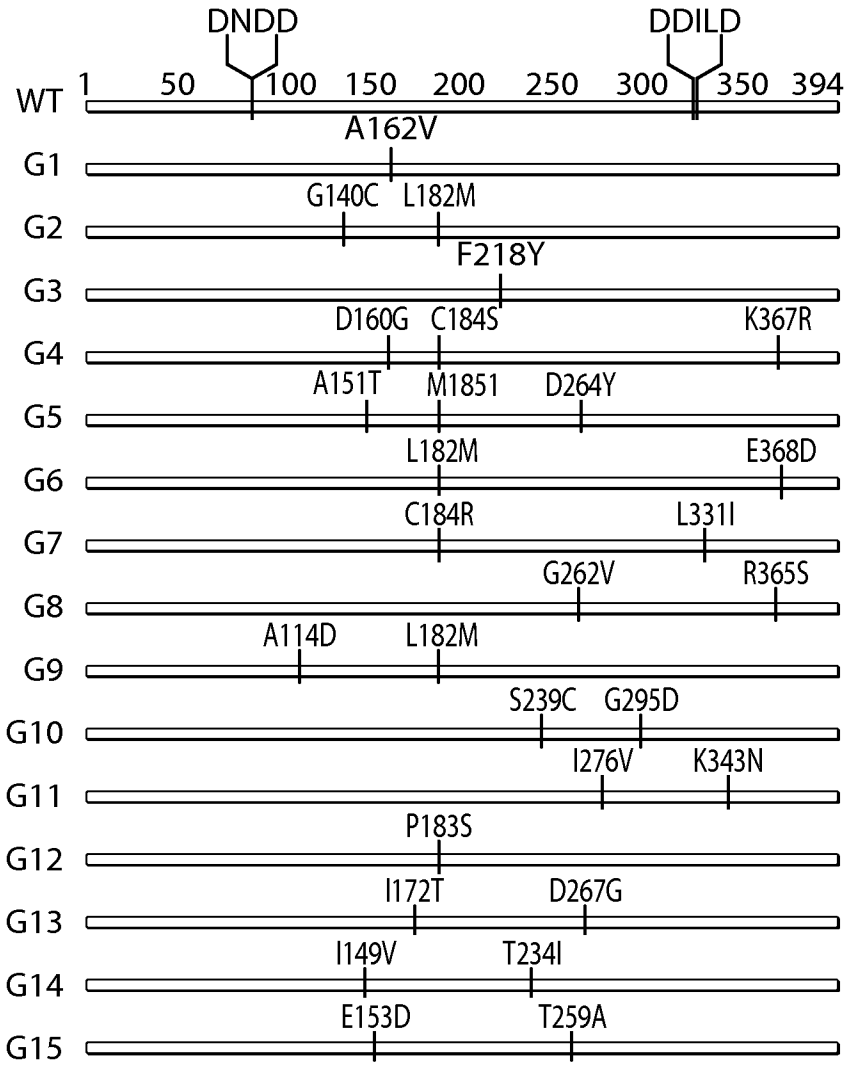

DNDD
DDILD
WT
1
50
100
150
200
250
300
350
394
A162V
G1
G140C
L182M
G2
F218Y
G3
D160G
C184S
K367R
G4
A151T
M1851
D264Y
G5
L182M
E368D
G6
C184R
L331I
G7
G262V
R365S
G8
A114D
L182M
G9
S239C
G295D
G10
I276V
K343N
G11
P183S
G12
I172T
D267G
G13
I149V
T234I
G14
E153D
T259A
G15

Fig. 4C

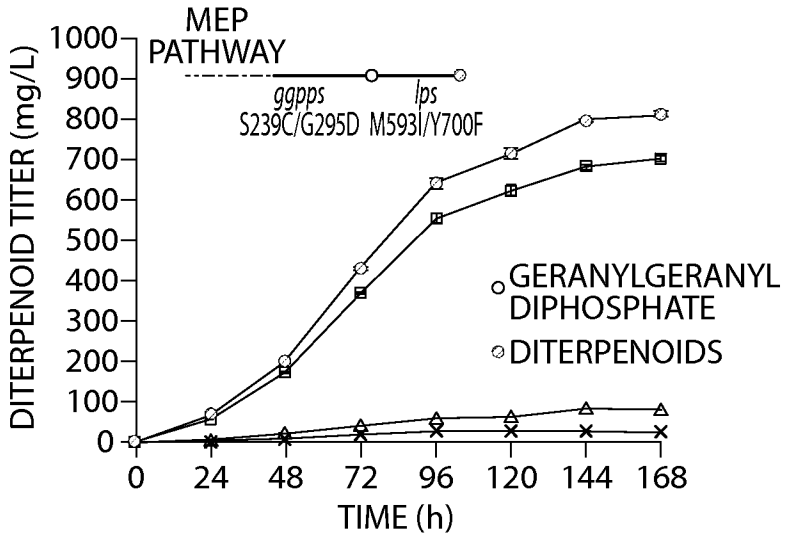

MEP
PATHWAY
ggpps
S239C/G295D
lps
M593I/Y700F
DITERPENOID TITER (mg/L)
1000
900
800
700
600
500
400
300
200
100
0
GERANYLGERANYL
DIPHOSPHATE
DITERPENOIDS
0
24
48
72
96
120
144
168
TIME (h)

Fig. 5A

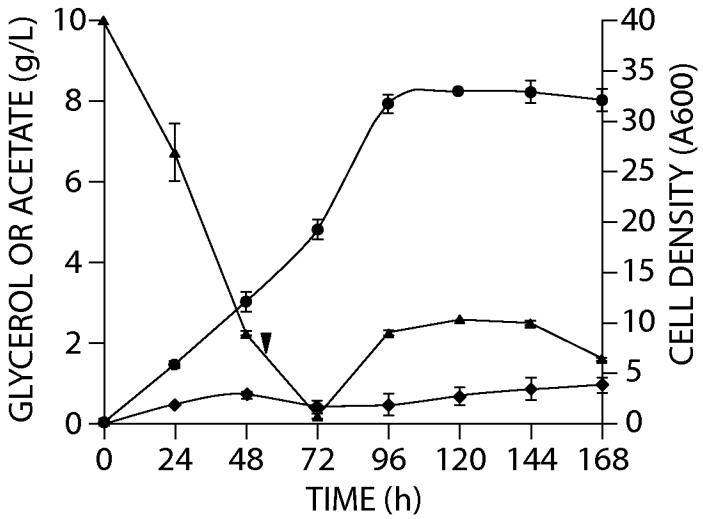

GLYCEROL OR ACETATE (g/L)
10
8
6
4
2
0
CELL DENSITY (A600)
40
35
30
25
20
15
10
5
0
0
24
48
72
96
120
144
168
TIME (h)

Fig. 5B

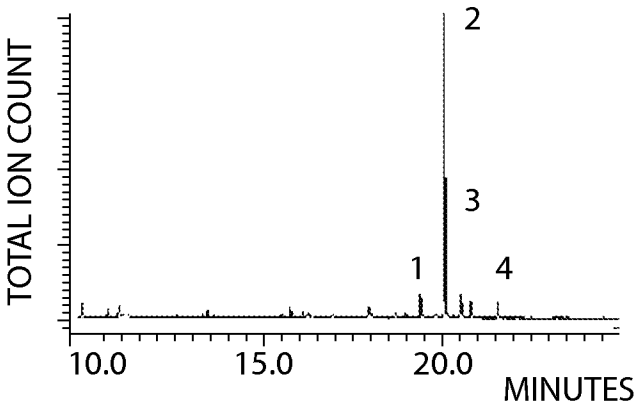


Fig. 6A

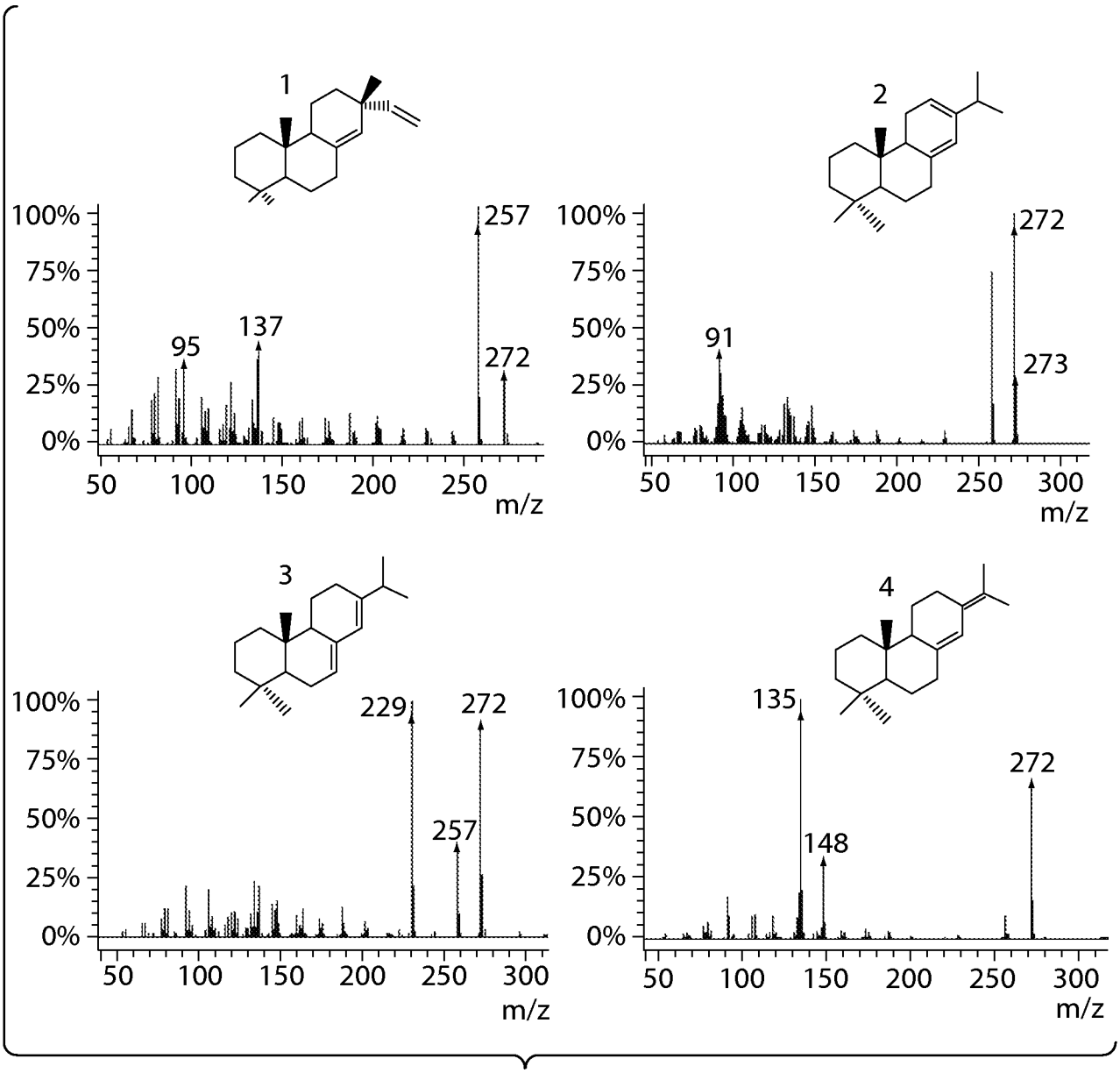

1
100%
75%
50%
25%
0%
257
137
95
272
50 100 150 200 250
m/z
2
100%
75%
50%
25%
0%
272
91
273
50 100 150 200 250 300
m/z
3
100%
75%
50%
25%
0%
229
272
257
50 100 150 200 250 300
m/z
4
100%
75%
50%
25%
0%
135
272
148
50 100 150 200 250 300
m/z

Fig. 6B

```
EAS   (1)--------------------------------------------------
LPS   (1)MAGVLFANLPCSLQLSPKVPFRQSTNILIPFHKRSSFGFNAQHCVRSHLR
EAS   (1)--------------------------------------------------
LPS  (51)LRWNCVGIHASAAETRPDQLPQEERFVSRLNADYHPAVWKDDFIDSLTSP
EAS   (1)--------------------------------------------------
LPS (101)NSHATSKSSVDETINKRIQTLVKEIQCMFQSMGDGETNPSAYDTAWVARI
EAS   (1)--------------------------------------------------
LPS (151)PSIDGSGAPQFPQTLQWILNNQLPDGSWGEECIFLAYDRVLNTLACLLTL
EAS   (1)--------------------------------------------------
LPS (201)KIWNKGDIQVQKGVEFVRKHMEEMKDEADNHRPSGFEVVFPAMLDEAKSL
EAS   (1)-----MASAAVANYEEEIVRPVADFSPSLWGDQFLSFSIDNQVAEKYIYA
LPS (251)GLDLPYHLPFISQIHQKRQKKLQKIPLNVLHNHQTALLYSLEGLQDVVDW
EAS  (46)QEIEALKEQTRSMLLATG----------RK--------------------
LPS (301)QEITNLQSRDGSFLSSPASTACVFMHTQNKRCLHFLNFVLSKFGDYVPCH
EAS  (66)----LADTLNLIDIIERLGISYHFEKEIDEILDQIYN---------QN--
LPS (351)YPLDLFERLWAVDTVERLGIDRYFKKEIKESLDYVYRYWDAERGVGWARC
EAS (101)SNCNDLCTSALQFRLLRQHGFNISPEIFSKFQDENG--K-FKESLASDVL
LPS (401)NPIPDVDDTAMGLRILRLHGYNVSSDVLENFRDEKGDFFCFAGQTQIGVT
EAS (148)GLLNLYEASHVRTHADDILEDALAFSTIHLESAAPH--------LKSPLR
LPS (451)DNLNLYRCSQVCFPGEKIMEEAKTFTTNHLQNALAKNNAFDKWAVKKDLP
EAS (190)EQVTHALEQCLHKGVPRVETRFFISSIYDKEQSK----------NNVLL
LPS (501)GEVEYAIKYPWHRSMPRLEARSYIEQFGSNDVWLGKTVYKMLYVSNEKYL
EAS (229)RFAKLDFNLLQMLHKQELAQVSRWWKDLDFVTTLPYARDRVVECYFWALG
LPS (551)ELAKLDFNMVQALHQKETQHIVSWWRESGFN-DLTFTRQRPVEMYFSVAV
EAS (279)VYFEPQYSQARVMLVKTISMISIVDDTFDAYGTVKELEAYTDAIQRWDIN
LPS (600)SMFEPEFAACRIAYAKTSCLAVILDDLYDTHGSLDDLKLFSEAVRRWDIS
EAS (329)EIDRLPD-YMKISYKAILDLYKDYEKELSSAGRSHIVCHAIERMKEVVRN
LPS (650)VLDSVRDNQLKVCFLGLYNTVNGFGKDGLKEQGRDVLGYLRKVWEGLLAS
EAS (378)YNVESTWFIEGYMPPVSEYLSNALATTTYYYLATTSYLG-MKSATEQDFE
LPS (700)YTKEAEWSAAKYVPTFNEYVENAKVSIALATVVLNSIFFTGELLPDYILQ
EAS (427)WLSKNPKILEASVIICRVIDDTATYEVEKSRGQIATGIECCMRDYG-IST
LPS (750)QVDLRSKFLHLVSLTGRLINDTKTYQAERNRGELVSSVQCYMRENPECTE
EAS (476)KEAMAKFQNMAETAWKDINEGLLRPTPVSTEFLTPILNLARIVEVTYIHN
LPS (800)EEALSHVYGIIDNALKELNWELANPASNAPLCVRRLLFNTARVMQLFYMY
EAS (526)LDGYTHPEKVLKPHIINLLVDSIKI
LPS (850)RDGFGISDKEMKDHVSRTLFDPVA-
```

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/056190

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P23/00 C12N1/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HUANG Q ET AL: "Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol", BIOORGANIC & MEDICINAL CHEMISTRY, PERGAMON, GB, vol. 9, no. 9, 1 September 2001 (2001-09-01), pages 2237-2242, XP002362540, ISSN: 0968-0896, DOI: DOI:10.1016/S0968-0896(01)00072-4	1-3, 11-16, 22-33, 35-37, 45-50, 56-63
Y	the whole document	18-21, 34, 52-55,64

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 January 2011	04/05/2011

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Luo, Xinmei

4

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	REILING K KINKEAD ET AL: "Mono and diterpene production in Escherichia coli", BIOTECHNOLOGY AND BIOENGINEERING, WILEY & SONS, HOBOKEN, NJ, US, vol. 87, no. 2, 20 July 2004 (2004-07-20), pages 200-212, XP009143370, ISSN: 0006-3592, DOI: DOI:10.1002/BIT.20128 [retrieved on 2004-06-18]	1-3, 11-16, 18, 22-33, 35-37, 45-50, 52,56-63
Y	the whole document	19-21, 34, 53-55,64

Y	WANG CHIA-WEI ET AL: "Directed evolution of metabolically engineered Escherichia coli for carotenoid production", BIOTECHNOLOGY PROGRESS, AMERICAN INSTITUTE OF CHEMICAL ENGINEERS, US, vol. 16, no. 6, 1 November 2000 (2000-11-01), pages 922-926, XP002524399, ISSN: 8756-7938, DOI: DOI:10.1021/BP000124F [retrieved on 2000-10-26] the whole document	18-21, 34, 52-55,64

X,P	LEONARD EFFENDI ET AL: "Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control.", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 3 AUG 2010 LNKD- PUBMED:20643967, vol. 107, no. 31, 3 August 2010 (2010-08-03), pages 13654-13659, XP002617176, ISSN: 1091-6490 the whole document	1-3, 11-37, 45-64

A	AJIKUMAR PARAYIL KUMARAN ET AL: "Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms.", MOLECULAR PHARMACEUTICS 2008 MAR-APR LNKD- PUBMED:18355030, vol. 5, no. 2, March 2008 (2008-03), pages 167-190, XP002617175, ISSN: 1543-8384 the whole document	1-3, 11-37, 45-64

	-/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/056190

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KIRBY JAMES ET AL: "Biosynthesis of plant isoprenoids: perspectives for microbial engineering", ANNUAL REVIEW OF PLANT BIOLOGY,, vol. 60, 1 January 2009 (2009-01-01), pages 335-355, XP009143317, the whole document -----	1-3, 11-37, 45-64
A	US 2002/164736 A1 (MATSUDA SEIICHI P T [US] ET AL) 7 November 2002 (2002-11-07) the whole document -----	1-3, 11-37, 45-64

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/056190

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   [ ] on paper

   [X] in electronic form

   b. (time)

   [X] in the international application as filed

   [ ] together with the international application in electronic form

   [ ] subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/056190

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 4-10, 38-44(completely); 1-3, 11-37, 45-64(partially)
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-64

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-64

   relates to a method comprising: recombinantly expressing a terpenoid synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme in a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway and the said cell;

   ---

2. claims: 65-77, 90-97

   relates to an isolated levopimaradiene synthase (LPS) polypeptide that contains a mutation at one or more of the residues selected from the list in claim 65 and a cell recombinantly expressing said mutant;

   ---

3. claims: 78-89, 98-104

   relates to a isolated geranylgeranyl diphosphate synthase (GGPPS) polypeptide, wherein the GGPPS polypeptide contains a mutation at residue S239 and/or G295 and a cell recombinantly expressing said mutant.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 4-10, 38-44(completely); 1-3, 11-37, 45-64(partially)

Present claim 1 relates to an extremely large number of possible methods. Support and disclosure in the sense of Article 6 and 5 PCT is to be found however for only a very small proportion of the methods claimed, see Examples in the description. The non-compliance with the substantive provisions is to such an extent, that the search was performed taking into consideration the non-compliance in determining the extent of the search of claim 1 (PCT Guidelines 9.19 and 9.23). The search of claim 1 was restricted to those claimed methods which appear to be supported, wherein the terpenoid synthase enzyme is levopimaradiene synthase enzyme and the cell is an Escherichia coli cell. Claims 4-10 relating to other cells were not searched.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2010/056190

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002164736 A1	07-11-2002	US 2007009996 A1	11-01-2007