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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 November 2013 (07.11.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/166320 A1

(51) International Patent Classification:
*C07K 14/415* (2006.01) *C12N 9/88* (2006.01)

(21) International Application Number:
PCT/US2013/039315

(22) International Filing Date:
2 May 2013 (02.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/641,861 2 May 2012 (02.05.2012) US
13/802,360 13 March 2013 (13.03.2013) US

(71) Applicants: **DANISCO US INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US). **THE GOODYEAR TIRE & RUBBER COMPANY** [US/US]; 1144 East Market Street, Akron, OH 44316-0001 (US).

(72) Inventors: **BEATTY, Mary, K.**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **HAYES, Kevin**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **HOU, Zhenglin**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **MEYER, David, J.**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **NANNAPANENI, Kishore**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **RIFE, Christopher, L.**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **WELLS, Derek, H.**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US). **ZASTROW-HAYES, Gina, M.**; 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).

(74) Agents: **SHIEH-NEWTON, Terri** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (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, BN, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, RW, SD, SL, SZ, TZ,

*[Continued on next page]*

(54) Title: LEGUME ISOPRENE SYNTHASE FOR PRODUCTION OF ISOPRENE

(57) Abstract: The present invention provides methods and compositions of polypeptides having isoprene synthase activity with improved performance characteristics. In particular, the present invention provides legume isoprene synthases for increased isoprene production in recombinant host cells.

Figure 1

peanut_fom_1n24. TRRSANYQPNLWDFEFLQSVENDLQVERLEERARKLEEEVRGLMKKVEIEPLSLLELMDN
Palba_from_1n24. ARRSANYEPNSWDYDYLLSSDTDESIEVYKDKAKKLEAEVRREINNEKAEFLTLLELIDN

peanut_fom_1n24. VERLGLTYKFEEDIKSALNNRIVPLLHHHTINKYG----LHATALSFRFLRQHAFHVSPD
Palba_from_1n24. VQRLGLGYRFESDIRGALDR-----FVSSGGFDAVTKTSLHGTALSFRLLRQHGFEVSQE

peanut_fom_1n24. VFESFKEEGK-FKKEISGDVLGLLNLYETSYLGFEGETILDEARAFSATHLKNLLQTNQV
Palba_from_1n24. AFSGFKDQNGNFLENLKEDIKAILSLYEASFLALEGENILDEAKVFAISHLKELSE-EKI

peanut_fom_1n24. QNKVMAEKVRHALELPYHRRVHRLEARWFIERYEQKEAHDGALLELAKLDFNMVQSVMKK
Palba_from_1n24. GKELAEQ-VNHALELPLHRRTQRLEAVWSIEAYRKKEDANQVLLELAILDYNMIQSVYQR

peanut_fom_1n24. ELQELSRWWREIGLTSKLDFVRDRLMEVYFWALGMAPHPQLTECRKAVTKMFGLVTIIDD
Palba_from_1n24. DLRETSRWWRRVGLATKLHFARDRLIESFYWAVGVAFEPQYSDCRNSVAKMFSFVTIIDD

peanut_fom_1n24. VYDVYGTLDELQLFTDAVDRWDVNAVETLPDYMKLCYLALYNSVNDTAYSTLREKGDNSL
Palba_from_1n24. IYDVYGTLDELELFTDAVERWDVNAINDLPDYMKLCFLALYNTINEIAYDNLKDKGENIL

peanut_fom_1n24. PHLAKSWRDLCKAFLQEAKWSNNKIIPPFDAYIRNASVSSSGGALLAPCYFSVTQDSTSQ
Palba_from_1n24. PYLTKAWADLCNAFLQEAKWLYNKSTPTFDDYFGNAWKSSSGPLQLVFAYFAVV-QNIKK

peanut_fom_1n24. -AIDSITNYHGIVRSSCAIFRLCNDLATSAAELERGETTNSITSYMTENGTTEEEARESL
Palba_from_1n24. EEIENLQKYHDTISRPSHIFRLCNDLASASAEIARGETANSVSCYMRTKGISEELATESV

peanut_fom_1n24. GKLIDQEWKKMNRDVVLESAYPNVFKEIAINMARVSHCTYQYGDGLGRPDDTAENRIKLS
Palba_from_1n24. MNLIDETWKKMNKEKLGGSLFAKPFVETAINLARQSHCTYHNGDAHTSPDELTRKRVLSV

peanut_fom_1n24. LIEPIP
Palba_from_1n24. ITEPIL

WO 2013/166320 A1

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, 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))*

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

# LEGUME ISOPRENE SYNTHASE FOR PRODUCTION OF ISOPRENE

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims priority to U.S. Provisional Patent Application No. 61/641,861, filed May 2, 2012, and U.S. Utility Patent Application No. 13/802,360, filed March 13, 2013, the disclosures of which are incorporated by reference herein in their entirety.

## INCORPORATION BY REFERENCE

**[0002]** The content of the following submission on ASCII text file (ST.25 text format) is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name is "643842004640_Sequence_Listing.txt"; date recorded: April 26, 2013; and the size of the ASCII text file in bytes is 185,972)

## FIELD OF THE INVENTION

**[0003]** The present invention provides methods and compositions comprising isoprene synthases for improved production of isoprene. In particular, the present invention provides legume isoprene synthases for increased isoprene production in host cells.

## BACKGROUND OF THE INVENTION

**[0004]** Isoprene (2-methyl-1,3-butadiene) is a volatile hydrocarbon that is insoluble in water and soluble in alcohol. Commercially viable quantities of isoprene can be obtained by direct isolation from petroleum C5 cracking fractions or by dehydration of C5 isoalkanes or isoalkenes (Weissermel and Arpe, Industrial Organic Chemistry, 4$^{th}$ ed., Wiley-VCH, pp. 117-122, 2003). The C5 skeleton can also be synthesized from smaller subunits. It would be desirable, however, to have a commercially viable method of producing isoprene that was independent of nonrenewable resources.

**[0005]** Isoprene monomer is employed in the manufacture of polyisoprene and various copolymers (with isobutylene, butadiene, styrene, or other monomers). Building a strain (prokaryotic or eukaryotic) capable of producing commercially viable levels of isoprene requires optimization of part of or the entire DXP or MVA pathway or both MVA and DXP pathways. A key enzyme in the pathway is isoprene synthase (IspS), which converts the precursor DMAPP to

isoprene. Isoprene synthases (IspS) that have been identified include those from plants such as poplar, English oak and kudzu vine. Some of the identified plant IspS enzymes have been partially characterized in part by expression in *E. coli* and some of the kinetic parameters of these enzymes have been determined *in vitro* with purified protein. However, kinetic parameters, including $K_m$, $K_i$, and $K_{cat}$, of the native IspS enzymes and even some of the recombinant enzymes are insufficient for commercial production of isoprene in a biological host. In addition, commercial use of some IspS enzymes can be limited due to insufficient expression levels. Thus, one problem to be solved is the provision of isoprene synthases that have improved properties such that a greater amount of isoprene can be biologically produced for commercial use. To solve this problem as described herein, an isoprene synthase with improved properties may be expressed in a host (e.g. a bacterial host). In particular, the present invention provides legume isoprene synthases that exhibit improved properties for increased isoprene production in host cells.

**[0006]** All patents, patent applications, publications, documents, nucleotide and protein sequence database accession numbers, the sequences to which they refer, and articles cited herein are all incorporated herein by reference in their entireties.

## BRIEF SUMMARY OF THE INVENTION

**[0007]** The invention provided herein discloses, *inter alia*, compositions of polypeptides having isoprene synthase activity with improved properties for production of isoprene and methods for identifying, producing and using such polypeptides for the production of isoprene. In some aspects, the present invention provides legume isoprene synthases for increased isoprene production in host cells.

**[0008]** Accordingly, in some aspects, the invention provides for an isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to SEQ ID NO: 1, wherein said polypeptide has one or more amino acid residue(s) corresponding to one or more amino acid residue(s) corresponding to SEQ ID NO:1, and wherein said one or more amino acid residue(s) are selected from the group consisting of F287, G397, N438, E451, and Y514. In other aspects, the isolated polypeptide is not from *P. montana*. In some aspects, the isolated polypeptide is an isoprene synthase selected from the group consisting of *Arachis sp.*, *Mucana sp.*, *Cajanus sp.*, *Glycine sp.*, *Lotus sp.*, and *Medicago*

*sp.* In further aspects, the isolated polypeptide is an isoprene synthase selected from the group consisting of *A. hypogaea*, *M. pruriens, C. cajans, G. max, G. soja, L. japonicus,* and *M. truncatula.* In yet a further aspect, the isolated polypeptide of claim 2, wherein the isolated polypeptide is an *A. hypogaea* isoprene synthase. In yet another aspect, the isolated polypeptide is an *M. pruriens* isoprene synthase. In some aspects, the invention provides legume isoprene synthases with at least one improved property as compared to a poplar isoprene synthase. In some aspects, at least one improved property is selected from but not limited to the group consisting of: specific productivity, yield, cellular performance index and protein expression. In some aspects, at least one improved property is selected from but not limited to the group consisting of: specific activity, $K_{cat}$, $K_i$, and $K_m$. In any of the aspects herein, the isolated polypeptide has a reduced substrate inhibition as compared to a poplar isoprene synthase. In any of the aspects herein, the isolated polypeptide has increased isoprene synthase activity as compared to a poplar isoprene synthase. In some aspects, the increased isoprene synthase activity is indicated by a host cell comprising the isoprene synthase displaying improved growth in the presence of mevalonic acid compared to a host cell comprising a poplar isoprene synthase. In any of the aspects herein, the isolated polypeptide is an isoprene synthase selected from the group consisting of: *A. hypogaea* (SEQ ID NO:3), *G. max* 1 (SEQ ID NO:5), *G. max* 2 (SEQ ID NO:7), *M. pruriens* (SEQ ID NO:9), and *C. cajans* (SEQ ID NO:11). In further aspects, the isolated polypeptide is an *A. hypogaea* isoprene synthase comprising the amino acid sequence of SEQ ID NO:3.

**[0009]** Additionally, in some aspects, the invention provides for an isolated polypeptide from a legume having isoprene synthase activity comprising at least 70% sequence identity to SEQ ID NO: 3. In any of the aspects herein, an isolated polypeptide from a legume having isoprene synthase activity, wherein said isoprene synthase has a $K_{cat}$ value of at least about 1.3 is provided. In any of the aspects herein, an isolated polypeptide from a legume having isoprene synthase activity, wherein said isoprene synthase has a $K_m$ value of at least about 2.5 is provided. In any of the aspects herein, an isolated polypeptide from a legume having isoprene synthase activity, wherein said isoprene synthase has a $K_{iDMAPP}$ value of at least about 13.0 is provided.

**[0010]** In addition, in some aspects, the invention provides for an isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to SEQ ID NO: 1, and wherein said polypeptide has a $K_{cat}$ value of at least

about 1.3. In other aspects, the invention provides for an isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to SEQ ID NO: 1, and wherein said polypeptide has a $K_m$ value of at least about 2.5. In yet other aspects, the invention provides for an isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to SEQ ID NO: 1, and wherein said polypeptide has a $K_{iDMAPP}$ value of at least about 13.0. In any of the aspects herein, a host cell comprising a heterologous polynucleotide sequence encoding an isoprene synthase in operable combination with a promoter is provided. In some aspects, the polynucleotide sequence is contained within a plasmid. In other aspects, the polynucleotide sequence is integrated into a chromosome of the host cell. In yet other aspects, the host is selected from the group consisting of gram-positive bacterial cells, gram-negative bacterial cells, filamentous fungal cells, and yeast cells. In other aspects, the host is selected from the group consisting of Escherichia sp. (*E. coli*), Panteoa sp. (*P. citrea*), Bacillus sp. (*B. subtilis*), Yarrowia sp. (*Y. lipolytica*), Trichoderma (*T. reesei*) and *Saccharomyces* (*S. cerevisiae*). In some aspects, the host cell is cultured in a medium comprising a carbon source selected from the group consisting of glucose, glycerol, glycerine, dihydroxyacetone, yeast extract, biomass, molasses, sucrose, and oil. In any of the aspects herein, the host cell further comprises one or more heterologous or native nucleic acid(s) encoding one or more mevalonate (MVA) pathway polypeptides. In some aspects, the host cell further comprises a heterologous nucleic acid encoding a mevalonate (MVA) pathway polypeptide selected from the group consisting of an MVA pathway polypeptide from *Saccharomyces cerevisiae*, *Enterococcus faecalis*, *E.casseliflavus* and *E.gallinarum*. In some other aspects, the host cell further comprising one or more nucleic acids endoding a heterologous or native nucleic acid encoding an IDI polypeptide is provided. In some aspects, the host cell further comprises a heterologous or native nucleic acid encoding an IDI polypeptide and/or a heterologous or native nucleic acid encoding a DXS polypeptide, optionally in combination with the native DXP pathway. In some aspects, the host cell further comprises one or more nucleic acids encoding an IDI polypeptide and the DXS polypeptide. In some further aspects, the host cell comprises one vector encoding the isoprene synthase, the IDI polypeptide, and the DXS polypeptide. In still other aspects, the host cell further comprises one or more nucleic acids encoding an MVA pathway polypeptide and/or a DXS polypeptide. In a further aspect, the host cell further comprises one or more nucleic acids encoding a DXS polypeptide, an IDI polypeptide, or one or more of the rest of the DXP pathway

polypeptides, and a MVA pathway polypeptide. In some further aspects, the host cell further comprising DXS is provided.

**[0011]** Additionally, in some aspects, the invention provides a method of producing isoprene, comprising: (a) culturing the host cells in any of the aspects herein under suitable culture conditions for production of isoprene; and (b) producing the isoprene. In some further aspects, the method further comprising (c) recovering the isoprene is provided. In yet some further aspects, the method further comprising (d) polymerizing isoprene is provided.

**[0012]** In addition, in some aspects, the invention provides a method of detecting isoprene synthase activity, comprising: (a) culturing a host cell comprising the expression vector under conditions suitable for producing a legume isoprene synthase; (b) lysing the host cells with a lysis buffer comprising lysozyme to produce a cell lysate; and (c) detecting isoprene synthase activity in the cell lysate by measuring isoprene production from dimethylallyl diphosphate (DMAPP). In some aspects, the host cell is selected from the group consisting of gram-positive bacterial cells, gram-negative bacterial cells, filamentous fungal cells, and yeast cells. In some further aspects, wherein the host cell is selected from the group consisting of *Escherichia sp. (E. coli), Panteoa sp. (P. citrea), Bacillus sp. (B. subtilis), Yarrowia sp. (Y. lipolytica), Trichoderma (T. reesei)*, and *Saccharomyces (S. cerevisiae)* is provided. In other aspects, the host cell is cultured in a medium that includes a carbon source selected from the group consisting of glucose, glycerol, glycerine, dihydroxyacetone, yeast extract, biomass, molasses, sucrose, and oil.

**[0013]** In some aspects, the invention provides a method of detecting isoprene in a plurality of samples in a high-throughput screening comprising: (a) providing: i) a plurality of samples each comprising isoprene synthase; ii) a glass plate comprising a plurality of wells; and iii) a seal for the glass plate; (b) placing the plurality of samples in the plurality of wells of the glass plate; (c) sealing the glass plate with the seal to produce a sealed glass plate having a headspace associated with the sample in each of the plurality of wells; (d) incubating the glass plate under conditions in which the isoprene synthase is active; and (e) detecting isoprene in the headspace. In one aspect, the isoprene is detected by gas chromatography-mass spectrometry (GC-MS). In further aspects, the method wherein the plurality of samples comprise host cells comprising an expression vector comprising a polynucleotide sequence encoding an isoprene synthase variant in operable combination with a promoter is provided. In other further aspects, the method

wherein the plurality of samples comprise a lysate of the host cells, lysozyme, and dimethylallyl diphosphate (DMAPP) is provided. In yet other further aspects, the method wherein the glass plate is a deep-well glass block is provided. In still yet other further aspects, the method wherein the plurality of wells comprises at least 24 wells is provided. In some further aspects, the method wherein the plurality of wells each comprise a volume of 2 ml or less is provided.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0014]** **Figure 1** is an amino acid sequence alignment between *A. hypogaea* isoprene synthase (also indicated as "peanut") and MEA *P. alba* isoprene synthase. These sequences are shown as starting from residue 3 (SEQ ID NOS: 56, 57).

**[0015]** **Figure 2A-F** is an amino acid sequence alignment of putative isoprene synthase enzymes with MEA *P. alba* isoprene synthase as a reference sequence. The consensus sequence is depicted at the bottom of the sequence alignment (SEQ ID NOS: 18, 20, 22, 24, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 26, 1, 28 respectively).

**[0016]** **Figure 3A-C** is an amino acid sequence alignment of putative isoprene synthase enzymes with *P. alba* isoprene synthase and *P. montana* isoprene synthase as reference sequences (SEQ ID NOS: 1, 15, 3, 58, 7, 5,59, 13, respectively).

**[0017]** **Figure 4** is a coomassie stained SDS-PAGE gel showing solubility of putative isoprene synthase enzymes as compared to a known isoprene synthase. **A)** shows protein levels of MEA *P. alba* isoprene synthase in the soluble fraction (lane 1) and in the insoluble fraction (lane 2). **B)** shows protein levels of *A. hypogaea* isoprene synthase in the total lysate (lane 2), soluble fraction (lane 3), and insoluble pellet fraction (lane 4); protein levels of *G. max 1* isoprene synthase in the total lysate (lane 5), soluble fraction (lane 6), and insoluble pellet fraction (lane 7); and protein levels of *G. max 2* isoprene synthase in the total lysate (lane 8), soluble fraction (lane 9), and insoluble pellet fraction (lane 10). Lane 1 contains a molecular weight marker. The arrow indicates a band corresponding to molecular weight of isoprene synthase.

**[0018]** **Figure 5** is a coomassie stained SDS-PAGE gel showing solubility of putative isoprene synthase enzymes as compared to a known isoprene synthase. Three dilutions of the soluble fraction were loaded in the lanes in order of decreasing protein concentration from left to

right. Protein levels of soluble *P. alba variant 3* isoprene synthase (lanes 2, 3, and 4), *M. pruriens* (lanes 5, 6, and 7), and *C. cajans* (lanes 8, 9, and 10) are shown. Lane 1 contains a molecular weight marker. The arrow indicates a band corresponding to molecular weight of isoprene synthase.

**[0019]** **Figure 6** is a graph demonstrating isoprene synthase enzymatic activity of putative isoprene synthases in an *in vitro* isoprene production assay. Supernatents prepared from lysates harvested from cells expressing a polypeptide isolated from *A. hypogaea* or *Q. petraea* were incubated with increasing concentrations of DMAPP and measured for raw isoprene production by GC-FID (Gas chromatograph-flame ionization detector).

**[0020]** **Figure 7** is a graph demonstrating isoprene synthase enzymatic activity of putative isoprene synthases in an *in vitro* isoprene production assay. Supernatents prepared from lysates harvested from cells expressing a polypeptide isolated from *M. pruriens* or an isoprene synthase isolated from *P. alba* variant 3 were incubated with increasing concentrations of DMAPP and measured for raw isoprene production by GC-FID (Gas chromatograph-flame ionization detector).

**[0021]** **Figure 8** is a graph demonstrating specific productivity of putative isoprene synthases as determined from an *in vitro* isoprene production assay. Supernatents prepared from lysates harvested from cells expressing a polypeptide isolated from *A. hypogaea*, *G. max 1*, *G. max 2*, *M. pruriens*, or *C. cajans* were incubated with increasing concentrations of DMAPP and measured for isoprene production by GC-FID. Specific productivity of isoprene synthase was calculated in units of mg isoprene/L/hr/OD. The poplar isoprene synthase from *P. alba* WT was used as a positive control.

**[0022]** **Figure 9** is a graph showing isoprene synthase specific activity of a polypeptide isolated from the legume *A. hypogaea* in an *in vitro* isoprene production assay. As compared to the isoprene synthase from *P. alba*, the *A. hypogaea* isoprene synthase exhibited reduced substrate inhibition at increasing concentrations of DMAPP. Isoprene synthase activity is reported in units of $s^{-1}$.

**[0023]** **Figure 10** is a graph showing isoprene synthase specific activity of a polypeptide isolated from the legume *A. hypogaea* in the presence of 10 mM or 100 mM magnesium (Mg) in

an *in vitro* isoprene production assay. The *A. hypogaea* isoprene synthase exhibited reduced substrate inhibition despite the presence of increasing DMAPP concentrations when a concentration of 100 mM $Mg^{2+}$ was provided. At a lower concentration of $Mg^{2+}$, the *A. hypogaea* isoprene synthase exhibited sensitivity to substrate inhibition at low concentrations of DMAPP. Isoprene synthase activity is reported in units of $s^{-1}$.

**[0024]** **Figure 11** is a graph showing isoprene synthase specific activity of a polypeptide isolated from the legume *A. hypogaea* in an *in vitro* isoprene production assay. As compared to the isoprene synthase from *P. alba WT*, the *A. hypogaea* isoprene synthase exhibited reduced substrate inhibition at increasing concentrations of DMAPP. The *A. hypogaea* isoprene synthase also exhibited reduced substrate inhibition at increasing concentrations of DMAPP when compared to variant isoprene synthases, *P. alba* variant 1 and *P. alba* variant 2, that had previously been determined to have increased isoprene synthase specific activity as compared to the parent *P. alba* sequence. Isoprene synthase activity is reported in units of $s^{-1}$.

**[0025]** **Figure 12** shows a plasmid map for pCL201

**[0026]** **Figure 13** is a graph showing the relationship between growth and DMAPP concentration in an assay strain. A bacterial strain expressing *A. hypogaea* (DW668) was grown in the presence of 0, 15, 30, 45, 60 or 75 mM mevalonate (MVA) 1hr after induction with IPTG and growth was measured by assaying optical density (OD) at 600 nm at the start of the assay and at every hour for 5hrs. Bacterial strains expressing an isoprene synthase from *P. alba* variant 1, *P. alba* variant 2, or *P. alba* variant 3 were used as a positive controls.

**[0027]** **Figure 14** is a graph showing the relationship between isoprene production and DMAPP concentration in an assay strain. A bacterial strain expressing *A. hypogaea* (DW668) was grown in the presence of 0, 15, 30, 45, 60 or 75 mM mevalonate (MVA) 1hr after induction with IPTG and isoprene production was measured by GC-FID at the start of the assay and at every hour for 5hrs. Specific productivity of isoprene synthase was calculated in units of mg isoprene/L/hr/OD. Bacterial strains expressing an isoprene synthase from *P. alba* variant 1, *P. alba* variant 2, or *P. alba* variant 3 were used as a positive controls.

**[0028]** **Figure 15** shows a plasmid map for pEWL1036

**[0029]** **Figure 16** is a graph showing the growth of a bacterial strain co-expressing a legume isoprene synthase with an upper MVA pathway polypeptides (*E. faecalis*) or upper MVA pathway polypeptides (*E.casseliflavus* and *E.gallinarum).* Growth was measured in bacterial strains co-expressing *A. hypogaea* with an *E. faecalis* upper MVA pathway polypeptides (EWL1043 and EWL1047), *E.casseliflavus* upper MVA pathway polypeptides (EWL1049), or *E.gallinarum* upper MVA pathway polypeptides (EWL1052) after induction with IPTG. Growth was assayed by optical density (OD) at 600 nm at the start of the growth assay and at every hour for 5hrs. Bacterial strains expressing an isoprene synthase from *P. alba* variant 1 with the different upper MVA pathways were used as a controls.

**[0030]** **Figure 17** is a graph showing the isoprene production of a bacterial strain co-expressing a legume isoprene synthase with an upper MVA pathway polypeptides (*E. faecalis*) or upper MVA pathway polypeptides (*E.casseliflavus* and *E.gallinarum).* Isoprene production was measured in bacterial strains co-expressing *A. hypogaea* with an *E. faecalis* upper MVA pathway polypeptides (EWL1043 and EWL1047), *E.casseliflavus* upper MVA pathway polypeptides (EWL1049), or *E.gallinarum* upper MVA pathway polypeptides (EWL1052) after induction with IPTG. Isoprene was measured by GC-FID at the start of the assay and at every hour for 5hrs. Specific productivity of isoprene synthase was calculated in units of mg isoprene/L/hr/OD. Bacterial strains expressing an isoprene synthase from *P. alba* variant 1 with the different upper MVA pathways were used as a controls.

## DETAILED DESCRIPTION

**[0031]** The present invention provides methods and compositions comprising at least one legume isoprene synthase enzyme with improved properties. The legume isoprene synthase comprises at least 40% sequence identity to SEQ ID NO: 1, wherein the legume isoprene synthase has one or more amino acid residue(s) corresponding to one or more amino acid residue(s) corresponding to SEQ ID NO: 1. In some aspects, one or more amino acid residue(s) can be F287, G397, N438, E451, and Y514. In some embodiments, the legume isoprene synthase is not from *P. montana.* The invention provides legume isoprene synthases with at least one improved property as compared to a poplar isoprene synthase. In some preferred embodiments, at least one improved property is selected from but not limited to the group consisting of: specific productivity, yield, cellular performance index and protein expression. In some particularly preferred embodiments, at least one improved property is selected from but

not limited to the group consisting of: specific activity, $K_{cat}$, $K_i$, and $K_m$. In particular, the present invention provides legume isoprene synthases for increased isoprene production in host cells. Biosynthetically produced isoprene of the present invention finds use in the manufacture of rubber, polymers, and elastomers.

**[0032]** Furthermore, the headings provided herein are not limitations of the various aspects or embodiments of the invention, which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole. Nonetheless, in order to facilitate understanding of the invention, a number of terms are defined below.

## I. General Techniques

**[0033]** The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, "*Molecular Cloning: A Laboratory Manual*", second edition (Sambrook et al., 1989); "*Oligonucleotide Synthesis*" (M. J. Gait, ed., 1984); "*Animal Cell Culture*" (R. I. Freshney, ed., 1987); "*Methods in Enzymology*" (Academic Press, Inc.); "*Current Protocols in Molecular Biology*" (F. M. Ausubel et al., eds., 1987, and periodic updates); "*PCR: The Polymerase Chain Reaction*", (Mullis et al., eds., 1994). Singleton et al., *"Dictionary of Microbiology and Molecular Biology"* 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), Baltz et al., "*Manual of Industrial Microbiology and Biotechnology*" 3rd ed., (Washington, DC : ASM Press, 2010), and March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4th ed., John Wiley & Sons (New York, N.Y. 1992), provide one skilled in the art with a general guide to many of the terms used in the present application.

## II. Definitions

**[0034]** Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although any methods and materials similar or equivalent to those described herein find use in the practice of the present invention, the preferred methods and materials are

described herein. Accordingly, the terms defined immediately below are more fully described by reference to the Specification as a whole.

**[0035]** As used herein, the term "polypeptides" includes polypeptides, proteins, peptides, fragments of polypeptides, and fusion polypeptides.

**[0036]** As used herein, an "isolated polypeptide" is not part of a library of polypeptides, such as a library of 2, 5, 10, 20, 50 or more different polypeptides and is separated from at least one component with which it occurs in nature. An isolated polypeptide can be obtained, for example, by expression of a recombinant nucleic acid encoding the polypeptide.

**[0037]** By "heterologous polypeptide" is meant a polypeptide encoded by a nucleic acid sequence derived from a different organism, species, or strain than the host cell. In some aspects, a heterologous polypeptide is not identical to a wild-type polypeptide that is found in the same host cell in nature. Examples of heterologous proteins include enzymes such as isoprene synthases. In some embodiments, the genes encoding the proteins are naturally occurring genes, while in other embodiments mutated and/or synthetic genes are used.

**[0038]** An "endogenous polypeptide" is a polypeptide whose amino acid sequence is naturally found in the host cell. In some embodiments, an endogenous polypeptide is identical to a wild-type polypeptide that is found in the host cell in nature.

**[0039]** As used herein, a "nucleic acid" refers to two or more deoxyribonucleotides and/or ribonucleotides covalently joined together in either single or double-stranded form. It is to be understood that mutations, including single nucleotide mutations, can occur within a nucleic acid as defined herein.

**[0040]** By "recombinant nucleic acid" is meant a nucleic acid of interest that is free of one or more nucleic acids (*e.g.*, genes) which, in the genome occurring in nature of the organism from which the nucleic acid of interest is derived, flank the nucleic acid of interest. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (*e.g.*, a cDNA, a genomic DNA fragment, or a cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. A recombinant nucleic acid may be obtained using molecular biology techniques

that are known in the art, or part or all of a recombinant nucleic acid may be chemically synthesized.

**[0041]** By "heterologous nucleic acid" is meant a nucleic acid sequence derived from a different organism, species or strain than the host cell. In some aspects, the heterologous nucleic acid is not identical to a wild-type nucleic acid that is found in the same host cell in nature. In some aspects, a heterologous nucleic acid is not identical to a wild-type nucleic acid that is found in the same host cell in nature.

**[0042]** An "endogenous nucleic acid" is a nucleic acid whose nucleic acid sequence is naturally found in the host cell. In some embodiments, an endogenous nucleic acid is identical to a wild-type nucleic acid that is found in the host cell in nature. In some embodiments, one or more copies of endogenous nucleic acids are introduced into a host cell.

**[0043]** A nucleic acid or protein of the invention may be in isolated or purified form. As used herein, "isolated," with respect to nucleic acid or protein, means separated from other components, such as, but not limited to a cell or cell culture. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques, such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein or nucleic acid that is the predominant species present in a preparation is substantially purified. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, "purified" means that when isolated, the isolate contains at least 90%, at least 95%, at least 98%, or more preferably at least 99% of nucleic acid or protein by weight of the isolate. Purified polypeptides may be obtained by a number of methods including, for example, laboratory synthesis, chromatography, preparative electrophoresis, gel electrophoresis, centrifugation, precipitation, affinity purification, etc. (see, generally, R Scopes, Protein Purification, Springer-Verlag, N.Y. (1982), Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990)).

**[0044]** As used herein, an "expression control sequence" means a nucleic acid sequence that directs transcription of a nucleic acid of interest. An expression control sequence can be a promoter, such as a constitutive or an inducible promoter, or an enhancer. An expression control sequence can be "native" or heterologous. A native expression control sequence is derived from

the same organism, species, or strain as the gene being expressed. A heterologous expression control sequence is derived from a different organism, species, or strain as the gene being expressed. An "inducible promoter" is a promoter that is active under environmental or developmental regulation.

**[0045]** By "operably linked" is meant a functional linkage between a nucleic acid expression control sequence (such as a promoter) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

**[0046]** The term "isoprene" refers to 2-methyl-1,3-butadiene (CAS# 78-79-5 ). It can be the direct and final volatile C5 hydrocarbon product from the elimination of pyrophosphate from 3,3-dimethylallyl diphosphate DMAPP. It may not involve the linking or polymerization of IPP molecules to DMAPP molecules. The term "isoprene" is not generally intended to be limited to its method of production unless indicated otherwise herein.

**[0047]** As used herein, the terms "isoprene synthase," "isoprene synthase variant", and "IspS," refer to enzymes that catalyze the elimination of pyrophosphate from diemethylallyl diphosphate (DMAPP) to form isoprene. An "isoprene synthase" may be a wild type sequence or an isoprene synthase variant.

**[0048]** As used herein, the term "naturally-occurring" refers to anything (e.g., proteins, amino acids, or nucleic acid sequences) that is found in nature (e.g., has not been manipulated by means of recombinant or chemical methods). As used herein, the term "non-naturally occurring" refers to anything that is not found in nature (e.g., recombinantly produced or chemically synthesized proteins, amino acids, or nucleic acid sequences produced in the laboratory).

**[0049]** As used herein, an amino acid residue of an amino acid sequence of interest that "corresponds to" or is "corresponding to" or in "correspondence with" an amino acid residue of a reference amino acid sequence indicates that the amino acid residue of the sequence of interest is at a location homologous or equivalent to an enumerated residue in the reference amino acid sequence. One skilled in the art can determine whether a particular amino acid residue position in a polypeptide corresponds to that of a homologous reference sequence. For example, the

sequence of an isoprene synthase polypeptide may be aligned with that of a reference sequence (e.g. SEQ ID NO: 1 using known techniques (e.g., basic local alignment search tool (BLAST), ClustalW2, Structure based sequences alignment program (STRAP), or the like). In addition, crystal structure coordinates of a reference sequence may be used as an aid in determining a homologous polypeptide residue's three dimensional structure (see, for example, PCT/US2010/032134). In another aspect, equivalent residues may be identified by determining homology at the level of tertiary structure. Using such methods, the amino acid residues of an isoprene synthase polypeptide or isoprene synthase variant may be numbered according to the corresponding amino acid residue position numbering of the reference sequence. For example, the amino acid sequence of SEQ ID NO: 1 may be used for determining amino acid residue position numbering of each amino acid residue of an isoprene synthase of interest.

**[0050]** As used herein, "homology" refers to sequence similarity or identity, with identity being preferred. Homology may be determined using standard techniques known in the art (see, e.g., Smith and Waterman, Adv. Appl. Math. 2:482 (1981); Needleman and Wunsch, J. Mol. Biol. 48:443 (1970); Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85:2444 (1988); software programs such as GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package (Genetics Computer Group, Madison, WI); and Devereux et al., Nucl. Acid Res. 12:387-395 (1984)). One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pair-wise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle (see Feng and Doolittle, J. Mol. Evol. 35:351-360 (1987)). The method is similar to that described by Higgins and Sharp (see Higgins and Sharp, CABIOS 5:151-153 (1989)). Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps. Another example of a useful algorithm is the BLAST algorithm, described by Altschul et al., (see Altschul et al., J. Mol. Biol. 215:403-410 (1990); and Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5787 (1993)). A particularly useful BLAST program is the WU-BLAST-2 program (see Altschul et al., Meth. Enzymol. 266:460-480 (1996)). WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and

composition of the particular database against which the sequence of interest is being searched. However, the values may be adjusted to increase sensitivity.

**[0051]** The percent sequence identity between a reference sequence and a test sequence of interest may be readily determined by one skilled in the art. The percent identity shared by polynucleotide or polypeptide sequences is determined by direct comparison of the sequence information between the molecules by aligning the sequences and determining the identity by methods known in the art. An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, (see Altschul, et al., J. Mol. Biol., 215:403-410 (1990)). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1992)) alignments (B) of 50, expectation (E) of 10, M'5, N'-4, and a comparison of both strands.

**[0052]** The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul, *supra*). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a isoprene synthase nucleic acid of this invention if the smallest sum probability in a comparison of the test nucleic acid to a isoprene synthase nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001. Where the test nucleic acid encodes an isoprene synthase polypeptide, it is considered similar to a specified isoprene synthase nucleic acid if the

comparison results in a smallest sum probability of less than about 0.5, and more preferably less than about 0.2.

**[0053]** Percent "identical" or "identity" in the context of two or more nucleic acid or polypeptide sequences refers to two or more sequences that are the same or have a specified percentage of nucleic acid residues or amino acid residues, respectively, that are the same, when compared and aligned for maximum similarity, as determined using a sequence comparison algorithm or by visual inspection. "Percent sequence identity" or "% identity" or "% sequence identity or "% amino acid sequence identity" of a subject amino acid sequence to a reference amino acid sequence means that the subject amino acid sequence is identical (i.e., on an amino acid-by-amino acid basis) by a specified percentage to the reference amino acid sequence over a comparison length when the sequences are optimally aligned. Thus, 80% amino acid sequence identity or 80% identity with respect to two amino acid sequences means that 80% of the amino acid residues in two optimally aligned amino acid sequences are identical.

**[0054]** "Percent sequence identity" or "% identity" or "% sequence identity" of a subject nucleic acid sequence to a reference nucleic acid sequence means that the subject nucleic acid sequence is identical (i.e., on a nucleotide-by-nucleotide basis for a polynucleotide sequence) by a specified percentage to the reference sequence over a comparison length when the sequences are optimally aligned. Thus, 80% nucleotide sequence identity or 80% identity with respect to two nucleic acid sequences means that 80% of the nucleotide residues in two optimally aligned nucleic acid sequences are identical.

**[0055]** The "percent sequence identity" or "% sequence identity" or "% identity" of a subject sequence to a reference sequence can be calculated by optimally aligning the two sequences and comparing the two optimally aligned sequences over the comparison length. The number of positions in the optimal alignment at which identical residues occur in both sequences is determined, thereby providing the number of matched positions, and the number of matched positions is then divided by the total number of positions of the comparison length (which, unless otherwise specified, is the length of the reference sequence). The resulting number is multiplied by 100 to yield the percent sequence identity of the subject sequence to the reference sequence.

**[0056]** "Optimal alignment" or "optimally aligned" refers to the alignment of two (or more) sequences giving the highest percent identity score. For example, optimal alignment of two polypeptide sequences can be achieved by manually aligning the sequences such that the maximum number of identical amino acid residues in each sequence are aligned together or by using software programs or procedures described herein or known in the art. Optimal alignment of two nucleic acid sequences can be achieved by manually aligning the sequences such that the maximum number of identical nucleotide residues in each sequence are aligned together or by using software programs or procedures described herein or known in the art.

**[0057]** Two sequences (e.g., polypeptide sequences) may be deemed "optimally aligned" when they are aligned using defined parameters, such as a defined amino acid substitution matrix, gap existence penalty (also termed gap open penalty), and gap extension penalty, so as to achieve the highest similarity score possible for that pair of sequences. The BLOSUM62 scoring matrix (see Henikoff and Henikoff, *supra*) is often used as a default scoring substitution matrix in polypeptide sequence alignment algorithms (e.g., BLASTP). The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each residue position in the gap. Exemplary alignment parameters employed are: BLOSUM62 scoring matrix, gap existence penalty=11, and gap extension penalty=1. The alignment score is defined by the amino acid positions of each sequence at which the alignment begins and ends (e.g., the alignment window), and optionally by the insertion of a gap or multiple gaps into one or both sequences, so as to achieve the highest possible similarity score.

**[0058]** Optimal alignment between two or more sequences can be determined manually by visual inspection or by using a computer, such as, but not limited to e.g., the BLASTP program for amino acid sequences and the BLASTN program for nucleic acid sequences (see, e.g., Altschul et al., Nucleic Acids Res. 25(17):3389-3402 (1997); see also the National Center for Biotechnology Information (NCBI) website) or CLUSTALW program.

**[0059]** The term "identical" in the context of two nucleic acids or polypeptide sequences refers to the residues in the two sequences that are the same when aligned for maximum correspondence, as measured using one of the following sequence comparison or analysis algorithms.

**[0060]** A polypeptide of interest may be said to be "substantially identical" to a reference polypeptide if the polypeptide of interest comprises an amino acid sequence having at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 99.5% sequence identity to the amino acid sequence of the reference polypeptide. The percent identity between two such polypeptides can be determined manually by inspection of the two optimally aligned polypeptide sequences or by using software programs or algorithms (e.g., BLAST, ALIGN, CLUSTAL) using standard parameters. One indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Typically, polypeptides that differ by conservative amino acid substitutions are immunologically cross-reactive. Thus, a polypeptide is substantially identical to a second polypeptide, e.g., where the two peptides differ only by a conservative amino acid substitution or one or more conservative amino acid substitutions.

**[0061]** A nucleic acid of interest may be said to be "substantially identical" to a reference nucleic acid if the nucleic acid of interest comprises a nucleotide sequence having at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at least about 99.5% sequence identity to the nucleotide sequence of the reference nucleic acid. The percent identity between two such nucleic acids can be determined manually by inspection of the two optimally aligned nucleic acid sequences or by using software programs or algorithms (e.g., BLAST, ALIGN, CLUSTAL) using standard parameters. One indication that two nucleic acid sequences are substantially identical is that the two nucleic acid molecules hybridize to each other under stringent conditions (e.g., within a range of medium to high stringency).

**[0062]** As used herein, the term "mass yield" refers to the mass of the product produced by the recombinant (*e.g.*, bacterial) cells divided by the mass of the glucose consumed by the recombinant cells multiplied by 100, or expressed as a percentage.

**[0063]** By "specific productivity," it is meant the mass of the product produced by the recombinant (bacterial) cell divided by the product of the time for production, the cell density, and the volume of the culture.

**[0064]** By "titer," it is meant the mass of the product produced by the recombinant (*e.g.*, bacterial) cells divided by the volume of the culture.

**[0065]** As used herein, the term "cell productivity index (CPI)" refers to the mass of the product produced by the recombinant (*e.g.,* bacterial) cells divided by the mass of the recombinant cells produced in the culture.

**[0066]** As used herein, the singular terms "a," "an," and "the" include the plural reference unless the context clearly indicates otherwise.

**[0067]** Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

**[0068]** Reference to "about" a value or parameter herein also includes (and describes) embodiments that are directed to that value or parameter per se.

**[0069]** It is understood that all aspects and embodiments of the invention described herein include "comprising," "consisting," and "consisting essentially of" aspects and embodiments. It is to be understood that methods or compositions "consisting essentially of" the recited elements include only the specified steps or materials and those that do not materially affect the basic and novel characteristics of those methods and compositions.

**[0070]** Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains.

**[0071]** It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

**[0072]** It is intended that every maximum numerical limitation given throughout this specification includes every lower numerical limitation, as if such lower numerical limitations were expressly written herein. Every minimum numerical limitation given throughout this specification will include every higher numerical limitation, as if such higher numerical limitations were expressly written herein. Every numerical range given throughout this specification will include every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein.

## III. Isoprene Synthase Sequences

**[0073]** Polypeptides can be identified as isoprene synthases by sequence alignment to a reference sequence, as well as by other characteristics (e.g., isoprene synthase activity). In some embodiments, the reference sequence is a plant isoprene synthase sequence. In some embodiments, the reference sequence is MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. In some embodiments, the reference sequence is MEA *P. alba* synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 2. In some embodiments, the reference sequence is *P. montana* (kudzu) isoprene synthase comprising the amino acid sequence of SEQ ID NO: 15. In some embodiments, the reference sequence is *P. montana* (kudzu) synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 16. In any aspects of the invention, sequence alignment to a reference sequence can be based on sequence alignment with mature polypeptide sequences, wherein said mature polypeptides are defined as having undergone and completed processing of immature signal sequence(s). In any aspects of the invention, sequence alignment to a reference sequence can be based on sequence alignment with the C-terminal region of the reference sequence, wherein the C-terminal region contains a catalytically active site. In aspects of the invention, sequence alignment of polypeptide sequences to a reference sequence will result in the identification of a consensus sequence, wherein said consensus sequence determines that the polypeptide has isoprene synthase activity. In one embodiment, the consensus sequence is shown in Figure 2A-F. In some embodiments, the reference sequence is a plant isoprene synthase sequence. In some embodiments, the reference sequence is MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. In some embodiments, the reference sequence is *P. montana* (kudzu) isoprene synthase comprising the amino acid sequence of SEQ ID NO: 15. In any aspects of the invention, identification of a polypeptide with isoprene synthase activity includes,

but is not limited to, said polypeptide having one or more amino acid residue(s) corresponding to one or more consensus poplar isoprene synthase amino acid residue(s). In some embodiments, the poplar isoprene synthase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, the one or more consensus poplar isoprene synthase residue(s) are selected from the group consisting of F287, G397, N438, E451, and Y514. In some aspects of the invention, consensus sequence refers to an archetypical amino acid sequence against which all candidate isoprene synthase sequences and sequence of interest are compared. In other aspects of the invention, consensus sequences refers to a sequence that sets forth the nucleotides that are most often present in a DNA sequence of interest. For each position of a gene, the consensus sequence gives the amino acid that is most abundant in that position in the multiple sequence alignment. In some aspects of the invention, multiple sequence alignment refers to the sequences of multiple homologs of a reference sequence that are aligned using an algorithm (*e.g.,* Clustal W). In some embodiments, the reference sequence is a plant isoprene synthase sequence. In some embodiments, the reference sequence is MEA *P. alba* synthase comprising the amino acid sequence of SEQ ID NO: 1. In some embodiments, the reference sequence is MEA *P. alba* synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 2. In some embodiments, the reference sequence is *P. montana* (kudzu) comprising the amino acid sequence of SEQ ID NO: 15. In some embodiments, the reference sequence is *P. montana* (kudzu) synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 16.

**[0074]** In some aspects of the invention, an isolated polypeptide from a legume having isoprene synthase comprises at least 40% sequence identity to SEQ ID NO: 1, wherein the legume isoprene synthase has one or more amino acid residue(s) corresponding to one or more amino acid residue(s) corresponding to SEQ ID NO: 1, wherein the one or more amino acid residue(s) are selected from the group consisting of F287, G397, N438, E451, and Y514. In some embodiments, the legume isoprene synthase is not from *P. montana.*

**[0075]** In any aspects of the invention, the polypeptide having isoprene activity is an isoprene synthase polypeptide. In some embodiments, the isoprene synthase polypeptide is from the family Leguminosae, the family Fagaceae, the family Rutaceae, the family Lauraceae, the family Vitaceae, or the family Myrtaceae. In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from a legume

including, but not limited to, peanut (such *Arachis hypogaea*), velvet bean (such as *Mucuna pruriens*), pigeon pea (such as *Cajanus cajans*), soybean (such as *Glycine max*), wild legume (such as *Lotus japonicus*), or clover-like plant (such as *Medicago truncatula*). In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from a citrus including, but not limited to, lemon (such *Citrus limon*). In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from *Cinnamomum tenuipile*. In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from a grape including, but not limited to, wine grape (such as *Vitis vinifera)*. In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from an oak such as *Querus petraea* or *Querus ilex*. In some embodiments, the isoprene synthase polypeptide or nucleic acid is a naturally-occurring polypeptide or nucleic acid from *Eucalyptus globulus* or *Melaleuca alternifolia*. In some embodiments, the isoprene synthase polypeptide or nucleic acid is not from *P. montana*. Isoprene synthases of the present invention include, but are not limited to, those identified by JGI Accession Nos. Glyma09g21900.1, Glyma20g18280.1, Glyma06g45780.1, and Glyma12g10990.1. Isoprene synthases of the present invention include, but are not limited to, those identified by a FGENESH predicted polypeptide encoded from a *G. max* genomic sequence, such as the polypeptide identified by Glyma20g18280_1_FG. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *A. hypogaea* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 3. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *A. hypogaea* isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 4.

**[0076]** In any aspects of the invention, the isoprene synthase polypeptide has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at

least about 99% sequence identity with MEA *P. alba* synthase comprising the amino acid sequence of SEQ ID NO: 1. In other embodiments, the isoprene synthase polypeptide has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with MEA *P. alba* synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 2. In any aspects of the invention, the isoprene synthase polypeptide has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *P. montana* (kudzu) comprising the amino acid sequence of SEQ ID NO: 15. In other embodiments, the isoprene synthase polypeptide has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *P. montana* (kudzu) synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 16.

**[0077]** Isoprene synthase variants may be generated from an isoprene synthase of the present invention, wherein the isoprene synthase may be an isoprene synthase as described herein, including wild type and non-wild type isoprene synthases. In some embodiments, the isoprene synthase is a legume isoprene synthase. Exemplary isoprene synthase nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of an isoprene synthase polypeptide. Exemplary isoprene synthase polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as variant polypeptides and nucleic acids derived from any of the source organisms described herein. In some embodiments, the isoprene synthase variants have one or more amino acid substitution(s) at one or more amino acid residues corresponding to one or more amino acid residue(s) corresponding to SEQ ID NO:1, wherein said one or more amino acid residue(s) corresponding to SEQ ID NO: 1 are selected from the group consisting of T481, K463, K393, L376, K161, S288, T240, E472, G389, and R242. In some embodiments, the amino acid substitution is S288C.

**[0078]** Several methods are known in the art that are suitable for generating variants of the isoprene synthases of the present invention, including but not limited to site-saturation

mutagenesis, scanning mutagenesis, insertional mutagenesis, random mutagenesis, site-directed mutagenesis, and directed-evolution, as well as various other recombinatorial approaches.

## IV. Isoprene Synthase Properties

**[0079]** In any aspects of the invention described herein, the polypeptide having isoprene synthase activity has at least one improved property over a reference sequence. In some embodiments, the reference sequence is a plant isoprene synthase sequence. In some embodiments, the reference sequence is MEA *P. alba* synthase comprising the amino acid sequence of SEQ ID NO: 1. In some embodiments, the reference sequence is MEA *P. alba* synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 2. In some embodiments, the reference sequence is *P. montana* (kudzu) comprising the amino acid sequence of SEQ ID NO: 15. In some embodiments, the reference sequence is *P. montana* (kudzu) synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 16. In some preferred embodiments, at least one improved property is selected from but not limited to the group consisting of: specific productivity, yield, cellular performance index and protein expression. In some particularly preferred embodiments, at least one improved property is selected from but not limited to the group consisting of: specific activity, $K_{cat}$, $K_i$, and $K_m$. In some embodiments, the polypeptide has a $K_{cat}$ value of at least about 1.3. In some embodiments, the polypeptide has a $K_m$ value of at least about 2.5. In some embodiments, the polypeptide has a $K_{iDMAPP}$ value of at least about 13.0.

**[0080]** Properties of interest include, but are not limited to: increased intracellular activity, specific productivity, yield, and cellular performance index. In some embodiments, specific productivity increase at least about 2, 3, 4, 5, 6 7, 8, 9, 10 times or more. In one embodiment, specific productivity is about 20 mg/L/OD/hr. In other embodiments, yield increase at least about 2, 3, 4, 5 times or more. In other embodiments, cell performance index increase at least about 2, 3, 4, 5 times or more. In other embodiments, intracellular activity increase at least about 2, 3, 4, 5, 6, 7, 8, 9, 10 times or more.

**[0081]** Without being bound by theory, these properties can be achieved by one or a combination of any of the following properties of IspS of the present invention: increased cellular viability, increased $k_{cat}$, decreased $K_m$, increased $K_m$, increased $K_i$, increased specific productivity, increased specific activity, increased solubility, decreased insolubility, improved

ribosome binding, increased translation initiation rate, increased translation elongation rate, increased transcription initiation rate, increased transcription elongation rate, decreased secondary structure of DNA, decreased secondary structure of RNA, increased secondary structure of DNA, increased secondary structure of RNA, increased folding rates, increased affinity for intracellular chaperones, increased stability, decreased protein turnover, increased protein expression levels, decreased exposure to intracellular protease, decreased affinity for intracellular protease, decreased localization to the periplasm, improved localization to the cytoplasm, decreased inclusion body formation, decreased membrane localization, increased expression due to a more favorable codon, increased DNA stability, increased RNA stability, and decreased RNA degradation. Any mutation that has a positive effect on the properties of nucleic acid sequences (DNA and RNA) encoding or expressing the IspS, or the biochemical properties of the IspS enzyme itself, could allow for greater activity within the cell. Other properties of interest include pH optima, temperature stability (e.g., $T_m$ value), as well as sensitivity to potential inhibitors including substrate or product inhibition. Oxidative and proteolytic stability are also of interest. Furthermore, activation or inhibition due to metal ion effects and ionic strength is of interest.

**[0082]** Growth index or performance index of a host cell comprising a nucleic acid encoding an isoprene synthase of the present invention may also be used to indicate whether a particular isoprene synthase has a property of interest. Growth index and performance index may be determined according to methods known to one of skill in the art and/or as taught herein. Growth and performance index may be determined for a particular isoprene synthase by comparison with a reference sequence. In some embodiments, the reference sequence is a plant isoprene synthase sequence. In some embodiments, the reference sequence is MEA *P. alba* synthase comprising the amino acid sequence of SEQ ID NO: 1. In some embodiments, the reference sequence is MEA *P. alba* synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 2. In some embodiments, the reference sequence is *P. montana* (kudzu) comprising the amino acid sequence of SEQ ID NO: 15. In some embodiments, the reference sequence is *P. montana* (kudzu) synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 16. In various embodiments, the growth index of the isoprene synthase is at least about 0.8, at least about 0.9, at least about 1.0, at least about 1.1, at least about 1.2, at least about 1.3, as compared with the reference sequence. In various embodiments, the performance index of the

isoprene synthase is at least about 0.7, at least about 0.8, at least about 0.9, at least about 1.0, at least about 1.1, at least about 1.2, as compared with the reference sequence.

**[0083]** Methods for determining the properties of interest are known tone of skill in the art. Certain methods are further described herein in the Examples. Polypeptides having isoprene synthase activity can be assessed based on the desired outcome or property to be improved. For example, an isoprene synthase can be tested for the conversion of DMAPP to isoprene *in vitro* with purified or partially purified isoprene synthase or *in vivo* in the context of a host organism such as *E. coli.* In some cases, the *E. coli* may also express the DXP pathway, the MVA pathway, or both. Improved activity is assessed in comparison with other isoprene synthases; for example, a *P. alba* isoprene synthase, a *P. alba* variant isoprene synthase, or other reference polypeptide. It is contemplated that enzymes having various degrees of e.g. stability, solubility, activity, and/or expression level in one or more of test conditions will find use in the present invention for the production of isoprene in a diversity of hosts. High throughput methods may provide an investigation of these properties in an economical manner.

**[0084]** There is a strong correlation between increased intracellular DMAPP levels and growth inhibition of *E. coli*, which can be alleviated by the expression of *P. alba* isoprene synthase (IspS). Without being bound by theory, increased levels of IspS activity should therefore allow for better growth due to more rapid conversion of DMAPP to isoprene. By monitoring the growth rates of *E. coli* expressing polypeptides of interest under these conditions, the inventors have identified IspS enzymes that display increased ability to convert DMAPP to isoprene within the cell. In some embodiments, the IspS is a legume IspS. In some embodiments, the IspS is *A. hypogaea* IspS. In other embodiments, the IspS is *M. pruriens* IspS.

**[0085]** The invention also contemplates methods for screening polypeptides for isoprene synthase activity, comprising: (a) contacting a host cell with a medium comprising about 10 μM to about 70 μM IPTG, and about 5 mM to about 20 mM mevalonic acid (MVA), wherein the host cell comprises a nucleic acid encoding a polypeptide of interest in operable combination with a promoter; and (b) measuring the growth rate of the host cell. The isoprene synthase growth rate may be compared to that of a reference isoprene synthase (e.g. *P. alba* WT isoprene synthase, MEA *P. alba* isoprene synthase, *P. alba* variant, or *P. montana*). The methods may be used to screen for polypeptides having isoprene synthase activity for a particular property of interest, for example, one or more of the properties described herein. In some embodiments, an

increased growth rate indicates an isoprene synthase with an increased ability to convert DMAPP to isoprene within the host cell synthase. Growth rates may be analyzed, for example, according to methods known in the art, or as exemplified in the Examples below. In some embodiments, the method further comprises determining a growth index for the isoprene synthase. In some embodiments, the method further comprises determining a performance index for the isoprene synthase. Growth rate of the cells in exponential phase and/or final density of the cells may be taken into consideration as factors when selecting isoprene synthases with improved properties.

**[0086]** In some embodiments, the IPTG is present in the medium at a concentration from about 10 μM to about 60 μM. In some embodiments, the IPTG is present in the medium at a concentration from about 20 μM to about 60 μM. In some embodiments, the IPTG is present in the medium at a concentration from about 40 μM to about 60 μM. In some embodiments, the IPTG is present in the medium at a concentration of about 50 μM. In some embodiments, the MVA is present in the medium at a concentration of about 5 mM to about 75 mM. In some embodiments, the MVA is present in the medium at a concentration of about 10 mM to about 60 mM. In some embodiments, the MVA is present in the medium at a concentration of about 15 mM to about 45 mM. In some embodiments, the MVA is present in the medium at a concentration of about 5 mM to about 30 mM. In some embodiments, the MVA is present in the medium at a concentration of about 5 mM to about 15 mM. In some embodiments, the MVA is present in the medium at a concentration of about 15 mM.

## V. Legume Isoprene Synthase for the Production of Isoprene

**[0087]** The invention features compositions and methods for the production of increased amounts of isoprene. In particular, these compositions and methods may increase the rate of isoprene production and the total amount of isoprene that is produced. The biosynthetic processes for isoprene production described herein are a desirable alternative to using natural rubber. As discussed further below, the amount of isoprene produced by cells can be greatly increased by introducing a heterologous nucleic acid encoding a isoprene synthase (IspS) of the present invention into the cells. In some embodiments, the isoprene synthase is a legume isoprene synthase.

*A. Recombinant cells for the production of isoprene*

**[0088]** Isoprene (2-methyl-1,3-butadiene) is an important organic compound used in a wide array of applications. For instance, isoprene is employed as an intermediate or a starting material in the synthesis of numerous chemical compositions and polymers, including in the production of synthetic rubber. Isoprene is also an important biological material that is synthesized naturally by many plants and animals.

**[0089]** Isoprene is produced from DMAPP by the enzymatic action of isoprene synthase. Therefore, without being bound to theory, it is thought that increasing the cellular production of isopentenyl pyrophosphate from mevalonate via the alternative lower MVA pathway in recombinant cells by any of the compositions and methods described above will likewise result in the production of higher amounts of isoprene. Increasing the molar yield of isopentenyl pyrophosphate production from glucose translates into higher molar yields of isoprene produced from glucose when combined with appropriate enzymatic activity levels of mevalonate kinase, phosphomevalonate decarboxylase, isopentenyl kinase, and other appropriate enzymes for isoprene production.

**[0090]** In some aspects, any of the cells used for the production of isoprene can express one or more heterologous nucleic acids encoding one or more polypeptides which increase the cellular production of isoprene. In other aspects, the cells can also harbor specific genomic mutations which either enhance the production of isoprene or which increase carbon availability through the metabolic pathways responsible for the production of isoprene (such as the MVA pathway).

**[0091]** As described herein, the present invention provides recombinant cells capable of producing of isoprene, wherein the cells comprise one or more nucleic acids encoding one or more polypeptides of the MVA pathway, and a heterologous nucleic acid encoding an isoprene synthase polypeptide, wherein culturing the cells in a suitable media provides for the production of isoprene. In a further embodiment, the recombinant cells further comprise one or more nucleic acids encoding an isopentenyl diphosphate isomerase (IDI) polypeptide. In certain embodiments, the present invention provides recombinant cells capable of isoprene production, wherein the cells comprise one or more nucleic acids encoding one or more polypeptides of the MVA pathway, and a heterologous nucleic acid encoding an isoprene synthase polypeptide,

wherein the cells produce increased amounts of isoprene compared to isoprene-producing cells that do not comprise a nucleic acid encoding a polypeptide having phosphomevalonate decarboxylase activity and/or a nucleic acid encoding a polypeptide having isopentenyl kinase activity. In a further embodiment, the recombinant cells further comprise one or more nucleic acids encoding an isopentenyl diphosphate isomerase (IDI) polypeptide.

**[0092]** Production of isoprene can also be made by using any of the recombinant host cells described herein further comprising one or more of the enzymatic pathways manipulations wherein enzyme activity is modulated to increase carbon flow towards mevalonate production. The recombinant cells described herein that have various enzymatic pathways manipulated for increased carbon flow to mevalonate production can be used to produce isoprene. In one embodiment, the recombinant cells further comprise a nucleic acid encoding a phosphoketolase. In another embodiment, the recombinant cells can be further engineered to incease the activity of one or more of the following genes selected from the group consisting of rribose-5-phosphate isomerase (*rpiA* and/or *rpiB*), D-ribulose-5-phosphate 3-epimerase (*rpe*), transketolase (*tktA* and/or *tktB*), transaldolase B (*tal B*), phosphate acetyltransferase (*pta* and/or *eutD*). In another embodiment, these recombinant cells can be further engineered to decrease the activity of one or more genes of the following genes including glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphofructokinase-1 (*pfkA* and/or *pfkB*), fructose bisphosphate aldolase (*fba*, *fbaA*, *fbaB*, and/or *fbaC*), glyceraldehyde-3-phosphate dehydrogenase (*gapA* and/or *gapB*), acetate kinase (*ackA*), citrate synthase (*gltA*), EI (*ptsI*), $EIICB^{Glc}$ (*ptsG*), $EIIA^{Glc}$ (*crr*), and/or HPr (*ptsH*).

## 1. Isoprene synthase nucleic acids and polypeptides

**[0093]** In some aspects of the invention, the recombinant cells described in any of the compositions or methods described herein express one or more nucleic acids encoding an isoprene synthase polypeptide or a polypeptide having isoprene synthase activity. In some aspects of the invention, the cells described in any of the compositions or methods described herein (including host cells that have been modified as described herein) further comprise one or more nucleic acids encoding an isoprene synthase polypeptide or a polypeptide having isoprene synthase activity. The nucleic acids and/or polypeptides can be either endogenous or heterologous. In some aspects, the isoprene synthase polypeptide is an endogenous polypeptide. In some aspects, the endogenous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a constitutive promoter. In some aspects, the endogenous nucleic acid

encoding an isoprene synthase polypeptide is operably linked to an inducible promoter. In some aspects, the endogenous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a strong promoter. In some aspects, more than one endogenous nucleic acid encoding an isoprene synthase polypeptide is used (e.g, 2, 3, 4, or more copies of an endogenous nucleic acid encoding an isoprene synthase polypeptide).

**[0094]** In some aspects, the isoprene synthase polypeptide is a heterologous polypeptide. In some aspects, the cells comprise more than one copy of a heterologous nucleic acid encoding an isoprene synthase polypeptide. In some aspects, the heterologous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a constitutive promoter. In some aspects, the heterologous nucleic acid encoding an isoprene synthase polypeptide is operably linked to an inducible promoter. In some aspects, the heterologous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a strong promoter. In some aspects, the heterologous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a weak promoter. In some aspects, the isoprene synthase polypeptide is a polypeptide or variant thereof from *Pueraria* or *Populus* or a hybrid such as *Populus alba x Populus tremula.* In some aspects, the isoprene synthase polypeptide is a polypeptide or variant thereof from *Pueraria montana* or *Pueraria lobata, Populus tremuloides, Populus alba, Populus nigra*, and *Populus trichocarpa.* In some aspects, the isoprene synthase polypeptide is from *Eucalyptus*.

**[0095]** The nucleic acids encoding an isoprene synthase polypeptide(s) can be integrated into a genome of the host cells or can be stably expressed in the cells. The nucleic acids encoding an isoprene synthase polypeptide(s) can additionally be on a vector.

**[0096]** In a particular aspect, the cells are engineered to overexpress the endogenous isoprene synthase pathway polypeptide relative to wild-type cells. In some aspects, the endogenous nucleic acid encoding an isoprene synthase polypeptide is operably linked to a weak promoter.

**[0097]** In some aspects, the isoprene synthase polypeptide is a polypeptide from the family Leguminosae, the family Fagaceae, the family Rutaceae, the family Lauraceae, the family Vitaceae, or the family Myrtaceae. In some embodiments, the isoprene synthase polypeptide is from a legume. In some embodiments, the legume isoprene synthase is *Arachis hypogaea* (SEQ ID: NO 3), *Glycine max* 1 (SEQ ID NO: 5), *Glycine max* 2 (SEQ ID NO: 7), *M. pruriens* (SEQ

ID NO: 9), *C. cajans* (SEQ ID NO: 11), *Lotus japonicas* (SEQ ID NO: 33), or *Medicago truncatula* (SEQ ID NO: 35). In some embodiments, the isoprene synthase polypeptide is from a citrus including, but not limited to, *Citrus limon* (SEQ ID NO: 17). In some embodiments, the isoprene synthase polypeptide is from *Cinnamomum tenuipile* (SEQ ID NO: 23). In some embodiments, the isoprene synthase polypeptide is from a grape including, but not limited to, *Vitis vinifera* (SEQ ID NO: 21). In some embodiments, the isoprene synthase polypeptide is from an oak such as *Querus petraea* (SEQ ID NO: 13) or *Querus ilex* (SEQ ID NO: 19). In some embodiments, the isoprene synthase polypeptide is from *Eucalyptus globulus* (SEQ ID NO: 25) or *Melaleuca alternifolia* (SEQ ID NO: 27). Isoprene synthase of the present invention include, but are not limited to, those identified by JGI Accession Nos. Glyma09g21900.1, Glyma20g18280.1, Glyma06g45780.1, and Glyma12g10990.1. Isoprene synthases of the present invention include, but are not limited to, those identified by a FGENESH predicted polypeptide encoded from a *G. max* genomic sequence, such as the polypeptide identified by Glyma20g18280_1_FG.

**[0098]** In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *A. hypogaea* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 3. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Glycine max* 1 isoprene synthase comprising the amino acid sequence of SEQ ID NO: 5. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Glycine max* 2 isoprene synthase comprising the amino acid sequence of SEQ ID NO: 7. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *M. pruriens* isoprene synthase

comprising the amino acid sequence of SEQ ID NO: 9. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *C. cajans* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 11. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Lotus japonicas* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 33. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Medicago truncatula* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 35.

**[0099]** In some aspects, the invention provides an isolated polypeptide having isoprene synthase activity, wherein the isolated polypeptide is not a legume. In some aspects, an isolated polypeptide from a citrus having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Citrus limon* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 17. In some aspects, an isolated polypeptide from a *Cinnamomum* having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Cinnamomum tenuipile* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 23. In some aspects, an isolated polypeptide from a grape having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Vitis vinifera* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 21. In some aspects, an isolated polypeptide from an oak having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least

about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Querus petraea* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 13. In some aspects, an isolated polypeptide from an oak having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Querus ilex* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 19. In some aspects, an isolated polypeptide from a eucalyptus having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Eucalyptus globulus* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 25. In some aspects, an isolated polypeptide from a manuka having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Melaleuca alternifolia* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 27.

**[0100]** The nucleic acids encoding an isoprene synthase polypeptide(s) can be integrated into a genome of the host cells or can be stably expressed in the cells. The nucleic acids encoding an isoprene synthase polypeptide(s) can additionally be on a vector (include multiple vectors).

**[0101]** Exemplary isoprene synthase nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of an isoprene synthase polypeptide. Isoprene synthase polypeptides convert dimethylallyl diphosphate (DMAPP) into isoprene. Exemplary isoprene synthase polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of an isoprene synthase polypeptide. Exemplary isoprene synthase polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein. In addition, variants of isoprene synthase can possess improved activity such as improved enzymatic activity. In some aspects, an isoprene synthase variant has other improved properties, such as improved stability (*e.g.*, thermo-stability), and/or improved solubility.

**[0102]** Standard methods can be used to determine whether a polypeptide has isoprene synthase polypeptide activity by measuring the ability of the polypeptide to convert DMAPP into isoprene *in vitro*, in a cell extract, or *in vivo*. Isoprene synthase polypeptide activity in the cell extract can be measured, for example, as described in Silver et al., J. Biol. Chem. 270:13010-13016, (1995). In one exemplary assay, DMAPP (Sigma) can be evaporated to dryness under a stream of nitrogen and rehydrated to a concentration of 100 mM in 100 mM potassium phosphate buffer pH 8.2 and stored at -20 0C. To perform the assay, a solution of 5 μL of 1M $MgCl_2$, 1 mM (250 μg/ml) DMAPP, 65 μL of Plant Extract Buffer (PEB) (50 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$, 5% glycerol, and 2 mM DTT) can be added to 25 μL of cell extract in a 20 ml Headspace vial with a metal screw cap and teflon coated silicon septum (Agilent Technologies) and cultured at 370C for 15 minutes with shaking. The reaction can be quenched by adding 200 μL of 250 mM EDTA and quantified by GC/MS.

**[0103]** In some aspects, the isoprene synthase polypeptide is a plant isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Pueraria* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from Populus or a variant thereof. In some aspects, the isoprene synthase polypeptide is a poplar isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is a kudzu isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is a willow isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is a eucalyptus isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is a polypeptide from *Pueraria* or *Populus* or a hybrid, *Populus alba x Populus tremula*, or a variant thereof. In some aspects, the isoprene synthase polypeptide is from *Robinia*, *Salix*, or *Melaleuca* or variants thereof.

**[0104]** In some embodiments, the plant isoprene synthase is from the family Fabaceae, the family Salicaceae, or the family Fagaceae. In some aspects, the isoprene synthase polypeptide or nucleic acid is a polypeptide or nucleic acid from *Pueraria montana* (kudzu) (Sharkey et al., *Plant Physiology* 137: 700-712, (2005)), *Pueraria lobata*, poplar (such as *Populus alba*, *Populus nigra*, *Populus trichocarpa*, or *Populus alba x tremula* (CAC35696) (Miller et al., Planta 213: 483-487, (2001)), aspen (such as *Populus tremuloides*) (Silver et al., JBC 270(22): 13010-1316, (1995)), English Oak (*Quercus robur*) (Zimmer et al., WO 98/02550), or a variant thereof. In

some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Pueraria montana*, *Pueraria lobata, Populus tremuloides*, *Populus alba*, *Populus nigra*, or *Populus trichocarpa or* a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Populus alba* or a variant thereof. In some aspects, the isoprene synthase is *Populus balsamifera* (Genbank JN173037), *Populus deltoides* (Genbank JN173039), *Populus fremontii* (Genbank JN173040), *Populus granidident*a (Genbank JN173038), Salix (Genbank JN173043), *Robinia pseudoacacia* (Genbank JN173041), *Wisteria* (Genbank JN173042), *Eucalyptus globulus* (Genbank AB266390) or *Melaleuca alterniflora* (Genbank AY279379) or variant thereof. In some aspects, the nucleic acid encoding the isoprene synthase (*e.g.*, isoprene synthase from *Populus alba* or a variant thereof) is codon optimized.

**[0105]** In some aspects, the isoprene synthase polypeptide is a legume isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Arachis* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Mucuna* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Cajanus* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Glycine* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Lotus* or a variant thereof. In some aspects, the isoprene synthase polypeptide is an isoprene synthase from *Medicago* or a variant thereof. In some aspects, the isoprene synthase polypeptide is a poplar isoprene synthase polypeptide or a variant thereof. In some aspects, the isoprene synthase polypeptide is a kudzu isoprene synthase polypeptide or a variant thereof.

**[0106]** In any aspects of the invention, the isoprene synthase polypeptide is encoded by a nucleic acid from the family Leguminosae, the family Fagaceae, the family Rutaceae, the family Lauraceae, the family Vitaceae, or the family Myrtaceae. In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from a legume. In some embodiments, the nucleic acid encoding legume isoprene synthase is from *Arachis hypogaea* (SEQ ID: NO 4), *Glycine max* 1 (SEQ ID NO: 6), *Glycine max* 2 (SEQ ID NO: 8), *M. pruriens* (SEQ ID NO: 10), *C. cajans* (SEQ ID NO: 12), wild legume (such as *Lotus japonicus*), or clover-like plant (such as *Medicago truncatula*). In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from a citrus including, but not limited to, lemon (such *Citrus limon*). In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from

*Cinnamomum tenuipile*. In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from a grape including, but not limited to, wine grape (such as *Vitis vinifera)*. In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from an oak such as *Querus petraea* (SEQ ID NO: 14) or *Querus ilex*. In some embodiments, the isoprene synthase polypeptide is encoded by a nucleic acid from *Eucalyptus globulus* or *Melaleuca alternifolia*. Isoprene synthase of the present invention include, but are not limited to, those identified by JGI Accession Nos. Glyma09g21900.1, Glyma20g18280.1, Glyma06g45780.1, and Glyma12g10990.1. Isoprene synthases of the present invention include, but are not limited to, those identified by a FGENESH predicted polypeptide encoded from a *G. max* genomic sequence, such as the polypeptide identified by Glyma20g18280_1_FG.

**[0107]** In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *A. hypogaea* isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 4. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Glycine max* 1 isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 6. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *Glycine max* 2 isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 8. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *M. pruriens* isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 10. In any aspects of the invention, an isolated polypeptide from a legume having isoprene synthase activity has at least about 40%, at least about 50%, at least about 60%, at least about 70%, at

least about 80%, at least about 90%, at least about 92%, at least about 94%, at least about 96%, at least about 98%, at least about 99% sequence identity with *C. cajans* isoprene synthase encoded by a nucleic acid sequence comprising the sequence of SEQ ID NO: 12.

**[0108]** In some aspects, the isoprene synthase nucleic acid or polypeptide is a naturally-occurring polypeptide or nucleic acid (*e.g.*, naturally-occurring polypeptide or nucleic acid from *Populus*). In some aspects, the isoprene synthase nucleic acid or polypeptide is not a wild-type or naturally-occurring polypeptide or nucleic acid. In some aspects, the isoprene synthase nucleic acid or polypeptide is a variant of a wild-type or naturally-occurring polypeptide or nucleic acid (*e.g.*, a variant of a wild-type or naturally-occurring polypeptide or nucleic acid from *Populus*).

**[0109]** In some aspects, the isoprene synthase nucleic acid or polypeptide is a naturally-occurring polypeptide or nucleic acid (e.g., naturally-occurring polypeptide or nucleic acid from *Arachis*). In some aspects, the isoprene synthase nucleic acid or polypeptide is not a wild-type or naturally-occurring polypeptide or nucleic acid. In some aspects, the isoprene synthase nucleic acid or polypeptide is a variant of a wild-type or naturally-occurring polypeptide or nucleic acid (*e.g.*, a variant of a wild-type or naturally-occurring polypeptide or nucleic acid from *Arachis*).

**[0110]** In some aspects, the isoprene synthase polypeptide is a legume isoprene synthase. In some aspects, the isoprene synthase polypeptide is a wild-type or naturally occurring legume isoprene synthase. In some aspects, the legume isoprene synthase has improved activity such as improved catalytic activity as compared to a reference isoprene synthase such MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. The increase in activity (*e.g.*, catalytic activity) can be at least about any of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 95%. In some aspects, the increase in activity such as catalytic activity is at least about any of 1 fold, 2 folds, 5 folds, 10 folds, 20 folds, 30 folds, 40 folds, 50 folds, 75 folds, or 100 folds. In some aspects, the increase in activity such as catalytic activity is about 10% to about 100 folds (*e.g.*, about 20% to about 100 folds, about 50% to about 50 folds, about 1 fold to about 25 folds, about 2 folds to about 20 folds, or about 5 folds to about 20 folds). In some aspects, the legume isoprene synthase has improved solubility compared to a reference isoprene synthase such MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1 The increase in solubility can be at least about any of 10%, 20%, 30%, 40%,

50%, 60%, 70%, 80%, 90%, or 95%. The increase in solubility can be at least about any of 1 fold, 2 folds, 5 folds, 10 folds, 20 folds, 30 folds, 40 folds, 50 folds, 75 folds, or 100 folds. In some aspects, the increase in solubility is about 10% to about 100 folds (*e.g.*, about 20% to about 100 folds, about 50% to about 50 folds, about 1 fold to about 25 folds, about 2 folds to about 20 folds, or about 5 folds to about 20 folds).

**[0111]** In some aspects, the $k_{cat}$ is at least about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0. In other aspects, the kcat is at least about 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6. 4.7, 4.8, 4.9, 5.0, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8. 5.9 or 6.0.

**[0112]** In some aspects, the $K_m$ is at least about 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 16, 17, 17.5, 18, 18.5, 19, 19.5, 20, 20.5, 21, 21.5, 22, 22.5, 23, 23.5, 24, 24.5 or 25.

**[0113]** In some aspects, $K_{iDMAPP}$ values at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60.

**[0114]** In some aspects, the legume isoprene synthase has at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100%, at least about 110%, at least about 120%, at least about 130%, at least about 140%, at least about 150%, at least about 160%, at least about 170%, at least about 180%, at least about 190%, at least about 200% to MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. The legume isoprene synthase can share sequence similarity MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. In some aspects, a legume isoprene synthase can have at least about any of 40%, 50%, 60%, 70%, 75%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% amino acid sequence identity as that of the MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1. In some aspects, a legume isoprene synthase has any of about 70% to about 99.9%, about 75% to about 99%, about 80% to about 98%, about 85% to about 97%, or about 90% to about 95% amino acid sequence identity as that of MEA *P. alba* isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1.

**[0115]** In some aspects, the isoprene synthase polypeptide is a variant. In some aspects, the isoprene synthase polypeptide is a variant of a wild-type or naturally occurring isoprene synthase. In some aspects, the variant has improved activity such as improved catalytic activity compared to the wild-type or naturally occurring isoprene synthase.

**[0116]** In any aspects of the invention, an isoprene synthase variant comprises a mutation in the wild-type or naturally occurring isoprene synthase. In some aspects, the isoprene synthase variant has at least one amino acid substitution, at least one amino acid insertion, and/or at least one amino acid deletion. In some aspects, the isoprene synthase variant has at least one amino acid substitution. In some aspects, the number of differing amino acid residues between the variant and wild-type or naturally occurring isoprene synthase can be one or more, *e.g.* 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, or more amino acid residues. Naturally occurring isoprene synthases can include any isoprene synthases from plants, for example, peanut (such *Arachis hypogaea*), velvet bean (such as *Mucuna pruriens*), pigeon pea (such as *Cajanus cajans*), soybean (such as *Glycine max*), wild legume (such as *Lotus japonicus*), clover-like plant (such as *Medicago truncatula*), citrus (such *Citrus limon*), *Cinnamomum tenuipile*, wine grape (such as *Vitis vinifera),* oak (such as *Querus petraea* or *Querus ilex), Eucalyptus globulus* or *Melaleuca alternifolia*. In some aspects, the variant is a variant of isoprene synthase from *Arachis hypogaea.* In some aspects, the variant of isoprene synthase from *Arachis hypogaea* has at least one amino acid substitution, at least one amino acid insertion, and/or at least one amino acid deletion. In some aspects, the variant is a truncated *Arachis hypogaea* isoprene synthase. In some aspects, the nucleic acid encoding variant (*e.g.*, variant of isoprene synthase from *Arachis hypogaea*) is codon optimized (for example, codon optimized based on host cells where the heterologous isoprene synthase is expressed). In some aspects, the variant is a variant of isoprene synthase from *Mucuna pruriens.* In some aspects, the variant of isoprene synthase from *Mucuna pruriens* has at least one amino acid substitution, at least one amino acid insertion, and/or at least one amino acid deletion. In some aspects, the variant is a truncated *Mucuna pruriens* isoprene synthase. In some aspects, the nucleic acid encoding variant (*e.g.*, variant of isoprene synthase from *Mucuna pruriens*) is codon optimized (for example, codon optimized based on host cells where the heterologous isoprene synthase is expressed).

**[0117]** The isoprene synthase polypeptide provided herein can be any of the isoprene synthases or isoprene synthase variants described in WO 2009/132220, WO 2010/124146, and

U.S. Patent Application Publication No.: 2010/0086978, the contents of which are expressly incorporated herein by reference in their entirety with respect to the isoprene synthases and isoprene synthase variants.

**[0118]** Any one of the promoters described herein (*e.g.*, promoters described herein and identified in the Examples of the present disclosure including inducible promoters and constitutive promoters) can be used to drive expression of any of the isoprene synthases described herein.

**[0119]** Suitable isoprene synthases include, but are not limited to, those identified by JGI Accession Nos. Glyma09g21900.1, Glyma20g18280.1, Glyma06g45780.1, and Glyma12g10990.1. Suitable isoprene synthases include, but are not limited to, those identified by a FGENESH predicted polypeptide encoded from a *G. max* genomic sequence, such as the polypeptide identified by Glyma20g18280_1_FG. Types of isoprene synthases which can be used in any one of the compositions or methods including methods of making microorganisms encoding isoprene synthase described herein are also described in International Patent Application Publication Nos. WO 2009/076676, WO 2010/003007, WO 2009/132220, WO 2010/031062, WO 2010/031068, WO 2010/031076, WO 2010/013077, WO 2010/031079, WO 2010/148150, WO 2010/124146, WO 2010/078457, and WO 2010/148256, U.S. Patent Application No. 13/283564, and Sharkey et al., "*Isoprene Synthase Genes Form A Monophyletic Clade Of Acyclic Terpene Synthases In The Tps-B Terpene Synthase Family*", Evolution (2012) (available on line at DOI: 10.1111/evo.12013), the contents of which are expressly incorporated herein by reference in their entirety with respect to the isoprene synthases and isoprene synthase variants.

## 2. MVA pathway nucleic acids and polypeptides

**[0120]** The complete MVA pathway can be subdivided into two groups: an upper and lower pathway. In the upper portion of the MVA pathway, acetyl Co-A produced during cellular metabolism is converted to mevalonate via the actions of polypeptides having either: (a) (i) thiolase activity or (ii) acetoacetyl-CoA synthase activity, (b) HMG-CoA reductase, and (c) HMG-CoA synthase enzymatic activity. First, acetyl Co-A is converted to acetoacetyl CoA via the action of a thiolase or an acetoacetyl-CoA synthase (which utilizes acetyl-CoA and malonyl-CoA). Next, acetoacetyl-CoA is converted to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by

the enzymatic action of HMG-CoA synthase. This Co-A derivative is reduced to mevalonate by HMG-CoA reductase, which is the rate-limiting step of the mevalonate pathway of isoprenoid production. In the lower MVA pathway, mevalonate is then converted into mevalonate-5-phosphate via the action of mevalonate kinase which is subsequently transformed into 5-diphosphomevalonate by the enzymatic activity of phosphomevalonate kinase. Finally, IPP is formed from 5-diphosphomevalonate by the activity of the enzyme mevalonate-5-pyrophosphate decarboxylase. The mevalonate-dependent biosynthetic pathway is particularly important for the production of the isoprenoid precursor molecules dimethylallyl diphosphate (DMAPP) and isopentenyl pyrophosphate (IPP).

**[0121]** Exemplary MVA pathway polypeptides include, but are not limited to: 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) polypeptides (*e.g.*, an enzyme encoded by *mvaS*), 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase) polypeptides (*e.g.*, enzyme encoded by *mvaR* or enzyme encoded by *mvaE* that has been modified to be thiolase-deficient but still retains its reductase activity), mevalonate kinase (MVK) polypeptides, phosphomevalonate kinase (PMK) polypeptides, diphosphomevalonte decarboxylase (MVD) polypeptides, phosphomevalonate decarboxylase (PMDC) polypeptides, isopentenyl phosphate kinase (IPK) polypeptides, IPP isomerase polypeptides, IDI polypeptides, and polypeptides (*e.g.*, fusion polypeptides) having an activity of two or more MVA pathway polypeptides. In particular, MVA pathway polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of an MVA pathway polypeptide. Exemplary MVA pathway nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of an MVA pathway polypeptide. Exemplary MVA pathway polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein. In addition, variants of MVA pathway polypeptide that confer the result of better isoprene production can also be used as well.

**[0122]** Non-limiting examples of MVA pathway polypeptides which can be used are described in International Patent Application Publication No. WO 2009/076676; WO 2010/003007 and WO 2010/148150.

a. Nucleic acids encoding polypeptides of the upper MVA pathway

**[0123]** The upper portion of the MVA pathway uses acetyl Co-A produced during cellular metabolism as the initial substrate for conversion to mevalonate via the actions of polypeptides having either: (a) (i) thiolase activity or (ii) acetoacetyl-CoA synthase activity, (b) HMG-CoA reductase, and (c) HMG-CoA synthase enzymatic activity. First, acetyl Co-A is converted to acetoacetyl CoA via the action of a thiolase or an acetoacetyl-CoA synthase (which utilizes acetyl-CoA and malonyl-CoA). Next, acetoacetyl-CoA is converted to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by the enzymatic action of HMG-CoA synthase. This Co-A derivative is reduced to mevalonate by HMG-CoA reductase, which is a rate-limiting step of the mevalonate pathway of isoprenoid production.

**[0124]** Non-limiting examples of upper MVA pathway polypeptides include: acetyl-CoA acetyltransferase (AA-CoA thiolase) polypeptides, acetoacetyl-CoA synthase polypeptides, 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) polypeptides, 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase) polypeptides. Upper MVA pathway polypeptides can include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of an upper MVA pathway polypeptide. Exemplary upper MVA pathway nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of an upper MVA pathway polypeptide. Exemplary MVA pathway polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein. Thus, it is contemplated herein that any gene encoding an upper MVA pathway polypeptide can be used in the present invention.

**[0125]** In certain embodiments, various options of *mvaE* and *mvaS* genes from *L. grayi, E. faecium, E. gallinarum, E. casseliflavus* and/or *E. faecalis* alone or in combination with one or more other *mvaE* and *mvaS* genes encoding proteins from the upper MVA pathway are contemplated within the scope of the invention. In other embodiments, an acetoacetyl-CoA synthase gene is contemplated within the scope of the present invention in combination with one or more other genes encoding: (i) 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) polypeptides and 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase) polypeptides. Thus, in certain aspects, any of the combinations of genes contemplated herein can be expressed in recombinant cells in any of the ways described herein.

**[0126]** Additional non-limiting examples of upper MVA pathway polypeptides which can be used herein are described in International Patent Application Publication No. WO 2009/076676; WO 2010/003007 and WO 2010/148150.

(i) Acetoacetyl-CoA synthase nucleic acids and polypeptides

**[0127]** The acetoacetyl-CoA synthase gene (aka *nphT7*) is a gene encoding an enzyme having the activity of synthesizing acetoacetyl-CoA from malonyl-CoA and acetyl-CoA and having minimal activity (*e.g.*, no activity) of synthesizing acetoacetyl-CoA from two acetyl-CoA molecules. *See, e.g.,* Okamura et al., PNAS Vol 107, No. 25, pp. 11265-11270 (2010), the contents of which are expressly incorporated herein for teaching about nphT7. An acetoacetyl-CoA synthase gene from an actinomycete of the genus Streptomyces CL190 strain was described in JP Patent Publication (Kokai) No. 2008-61506 A and US 2010/0285549. Acetoacetyl-CoA synthase can also be referred to as acetyl CoA:malonyl CoA acyltransferase. A representative acetoacetyl-CoA synthase (or acetyl CoA:malonyl CoA acyltransferase) that can be used is Genbank AB540131.1.

**[0128]** In one embodiment, acetoacetyl-CoA synthase of the present invention synthesizes acetoacetyl-CoA from malonyl-CoA and acetyl-CoA via an irreversible reaction. The use of acetoacetyl-CoA synthase to generate acetyl-CoA provides an additional advantage in that this reaction is irreversible while acetoacetyl-CoA thiolase enzyme's action of synthesizing acetoacetyl-CoA from two acetyl-CoA molecules is reversible. Consequently, the use of acetoacetyl-CoA synthase to synthesize acetoacetyl-CoA from malonyl-CoA and acetyl-CoA can result in significant improvement in productivity for isoprene compared with using thiolase to generate the end same product.

**[0129]** Furthermore, the use of acetoacetyl-CoA synthase to produce isoprene provides another advantage in that acetoacetyl-CoA synthase can convert malonyl CoA to acetyl CoA via decarboxylation of the malonyl CoA. Thus, stores of starting substrate are not limited by the starting amounts of acetyl CoA. The synthesis of acetoacetyl-CoA by acetoacetyl-CoA synthase can still occur when the starting substrate is only malonyl-CoA. In one embodiment, the pool of starting malonyl-CoA is increased by using host strains that have more malonyl-CoA. Such increased pools can be naturally occurring or be engineered by molecular manipulation. See, for

example Fowler, et al., Applied and Environmental Microbiology, Vol. 75, No. 18, pp. 5831-5839 (2009).

**[0130]** In any of the aspects or embodiments described herein, an enzyme that has the ability to synthesize acetoacetyl-CoA from malonyl-CoA and acetyl-CoA can be used. Non-limiting examples of such an enzyme are described herein. In certain embodiments described herein, an acetoacetyl-CoA synthase gene derived from an actinomycete of the genus Streptomyces having the activity of synthesizing acetoacetyl-CoA from malonyl-CoA and acetyl-CoA can be used. An example of such an acetoacetyl-CoA synthase gene is the gene encoding a protein having the amino acid sequence of SEQ ID NO: 15. Such a protein having the amino acid sequence of SEQ ID NO: 15 corresponds to an acetoacetyl-CoA synthase having activity of synthesizing acetoacetyl-CoA from malonyl-CoA and acetyl-CoA and having no activity of synthesizing acetoacetyl-CoA from two acetyl-CoA molecules.

*Sequence of acetoacetyl-CoA synthase*

**[0131]** MTDVRFRIIGTGAYVPERIVSNDEVGAPAGVDDDWITRKTGIRQ RRWAADDQATSDLATAAGRAALKAAGITPEQLTVIAVATSTPDRPQPPTAAYVQHHLG ATGTAAFDVNAVCSGTVFALSSVAGTLVYRGGYALVIGADLYSRILNPADRKTVVLFG DGAGAMVLGPTSTGTGPIVRRVALHTFGGLTDLIRVPAGGSRQPLDTDGLDAGLQYFA MDGREVRRFVTEHLPQLIKGFLHEAGVDAADISHFVPHQANGVMLDEVFGELHLPRAT MHRTVETYGNTGAASIPITMDAAVRAGSFRPGELVLLAGFGGGMAASFALIEW **(SEQ ID NO: 15)**.

**[0132]** The acetoacetyl-CoA synthase activity of a polypeptide can be evaluated as described below. Specifically, a gene encoding a polypeptide to be evaluated is first introduced into a host cell such that the gene can be expressed therein, followed by purification of the protein by a technique such as chromatography. Malonyl-CoA and acetyl-CoA are added as substrates to a buffer containing the obtained protein to be evaluated, followed by, for example, incubation at a desired temperature (e.g., 10°C to 60°C). After the completion of reaction, the amount of substrate lost and/or the amount of product (acetoacetyl-CoA) produced are determined. Thus, it is possible to evaluate whether or not the protein being tested has the function of synthesizing acetoacetyl-CoA from malonyl-CoA and acetyl-CoA and to evaluate the degree of synthesis. In such case, it is possible to examine whether or not the protein has the activity of synthesizing

acetoacetyl-CoA from two acetyl-CoA molecules by adding acetyl-CoA alone as a substrate to a buffer containing the obtained protein to be evaluated and determining the amount of substrate lost and/or the amount of product produced in a similar manner.

(ii) Genes encoding mvaE and mvaS polypeptides

**[0133]** In certain embodiments, various options of *mvaE* and *mvaS* genes from *L. grayi, E. faecium, E. gallinarum, E. casseliflavus* and/or *E. faecalis* alone or in combination with one or more other *mvaE* and *mvaS* genes encoding proteins from the upper MVA pathway are contemplated within the scope of the invention. In some embodiments, the *mvaE* gene encodes a polypeptide that possesses both thiolase and HMG-CoA reductase activities (Hedl, et al., J Bacteriol. 2002 April; 184(8): 2116–2122). In *L. grayi, E. faecium, E. gallinarum, E. casseliflavus,* and *E. faecalis*, the *mvaE* gene encodes a polypeptide that possesses both thiolase and HMG-CoA reductase activities. In fact, the *mvaE* gene product represented the first bifunctional enzyme of IPP biosynthesis found in eubacteria and the first example of HMG-CoA reductase fused to another protein in nature (Hedl, et al., J. Bacteriol. 2002 April; 184(8): 2116–2122). The *mvaS* gene, on the other hand, can encode a polypeptide having an HMG-CoA synthase activity.

**[0134]** Accordingly, recombinant cells (*e.g.*, *E. coli*) can be engineered to express one or more *mvaE* and *mvaS* genes from *L. grayi, E. faecium, E. gallinarum, E. casseliflavus* and/or *E. faecalis* to produce isoprene. The one or more *mvaE* and *mvaS* genes can be expressed on a multicopy plasmid. The plasmid can be a high copy plasmid, a low copy plasmid, or a medium copy plasmid. Alternatively, the one or more *mvaE* and *mvaS* genes can be integrated into the host cell's chromosome. For both heterologous expression of the one or more *mvaE* and *mvaS* genes on a plasmid or as an integrated part of the host cell's chromosome, expression of the genes can be driven by either an inducible promoter or a constitutively expressing promoter. The promoter can be a strong driver of expression, it can be a weak driver of expression, or it can be a medium driver of expression of the one or more *mvaE* and *mvaS* genes.

**[0135]** The *mvaE* gene encodes a polypeptide that possesses both thiolase and HMG-CoA reductase activities. The thiolase activity of the polypeptide encoded by the *mvaE* gene converts acetyl Co-A to acetoacetyl CoA whereas the HMG-CoA reductase enzymatic activity of the polypeptide converts 3-hydroxy-3-methylglutaryl-CoA to mevalonate. Exemplary mvaE

polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein that have at least one activity of a mvaE polypeptide.

**[0136]** Mutant mvaE polypeptides include those in which one or more amino acid residues have undergone an amino acid substitution while retaining mvaE polypeptide activity (*i.e.*, the ability to convert acetyl Co-A to acetoacetyl CoA as well as the ability to convert 3-hydroxy-3-methylglutaryl-CoA to mevalonate). The amino acid substitutions can be conservative or non-conservative and such substituted amino acid residues can or cannot be one encoded by the genetic code. The standard twenty amino acid "alphabet" has been divided into chemical families based on similarity of their side chains. Those families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a chemically similar side chain (*i.e.*, replacing an amino acid having a basic side chain with another amino acid having a basic side chain). A "non-conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a chemically different side chain (*i.e.*, replacing an amino acid having a basic side chain with another amino acid having an aromatic side chain).

**[0137]** Amino acid substitutions in the mvaE polypeptide can be introduced to improve the functionality of the molecule. For example, amino acid substitutions that increase the binding affinity of the mvaE polypeptide for its substrate, or that improve its ability to convert acetyl Co-A to acetoacetyl CoA and/or the ability to convert 3-hydroxy-3-methylglutaryl-CoA to mevalonate can be introduced into the mvaE polypeptide. In some aspects, the mutant mvaE polypeptides contain one or more conservative amino acid substitutions.

**[0138]** In one aspect, mvaE proteins that are not degraded or less prone to degradation can be used for the production of isoprene. Examples of gene products of *mvaEs* that are not degraded or less prone to degradation which can be used include, but are not limited to, those

from the organisms *E. faecium*, *E. gallinarum*, *E. casseliflavus*, *E. faecalis,* and *L.* grayi. One of skill in the art can express mvaE protein in *E. coli* BL21 (DE3) and look for absence of fragments by any standard molecular biology techniques. For example, absence of fragments can be identified on Safestain stained SDS-PAGE gels following His-tag mediated purification or when expressed in mevalonate or isoprene producing *E. coli* BL21 using the methods of detection described herein.

**[0139]** Standard methods, such as those described in Hedl et al., (J. Bacteriol. 2002, April; 184(8): 2116–2122) can be used to determine whether a polypeptide has mvaE activity, by measuring acetoacetyl-CoA thiolase as well as HMG-CoA reductase activity. In an exemplary assay, acetoacetyl-CoA thiolase activity is measured by spectrophotometer to monitor the change in absorbance at 302 nm that accompanies the formation or thiolysis of acetoacetyl-CoA. Standard assay conditions for each reaction to determine synthesis of acetoacetyl-CoA, are 1 mM acetyl-CoA, 10 mM $MgCl_2$, 50 mM Tris, pH 10.5 and the reaction is initiated by addition of enzyme. Assays can employ a final volume of 200 µl. For the assay, 1 enzyme unit (eu) represents the synthesis or thiolysis in 1 min of 1 µmol of acetoacetyl-CoA. In another exemplary assay, of HMG-CoA reductase activity can be monitored by spectrophotometer by the appearance or disappearance of NADP(H) at 340 nm. Standard assay conditions for each reaction measured to show reductive deacylation of HMG-CoA to mevalonate are 0.4 mM NADPH, 1.0 mM (*R*,*S*)-HMG-CoA, 100 mM KCl, and 100 mM $K_xPO_4$, pH 6.5. Assays employ a final volume of 200 µl. Reactions are initiated by adding the enzyme. For the assay, 1 eu represents the turnover, in 1 min, of 1 µmol of NADP(H). This corresponds to the turnover of 0.5 µmol of HMG-CoA or mevalonate.

**[0140]** Exemplary mvaE nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of a mvaE polypeptide. Exemplary mvaE polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein. Exemplary mvaE nucleic acids include, for example, mvaE nucleic acids isolated from *Listeria grayi_DSM 20601,* Enterococcus *faecium, Enterococcus gallinarum EG2, Enterococcus faecalis,* and/or *Enterococcus casseliflavus.* The mvaE nucleic acid encoded by the *Listeria grayi_DSM 20601 mvaE* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%,

92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:47. The mvaE nucleic acid encoded by the *Enterococcus faecium mvaE* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:48. The mvaE nucleic acid encoded by the *Enterococcus gallinarum EG2 mvaE* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:49. The mvaE nucleic acid encoded by the *Enterococcus casseliflavus mvaE* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:50. The mvaE nucleic acid encoded by the *Enterococcus faecalis mvaE* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to the *mvaE* gene previously disclosed in *E. coli* to produce mevalonate (see US 2005/0287655 A1; Tabata, K. and Hashimoto,S.-I. Biotechnology Letters 26: 1487–1491, 2004).

*Sequence of Listeria grayi DSM 20601 mvaE*

atggttaaagacattgtaataattgatgccctccgtactcccatcggtaagtaccgcggtcagctctcaaagatgacggcggtggaattggga
accgcagttacaaaggctctgttcgagaagaacgaccaggtcaaagaccatgtagaacaagtcatttttggcaacgttttacaggcaggga
acggccagaatcccgcccgtcagatcgcccttaattctggcctgtccgcagagataccggcttcgactattaaccaggtgtgtggttctggc
ctgaaagcaataagcatggcgcgccaacagatcctactcggagaagcggaagtaatagtagcaggaggtatcgaatccatgacgaatgc
gccgagtattacatattataataaagaagaagacaccctctcaaagcctgttcctacgatgaccttcgatggtctgaccgacgcgtttagcgg
aaagattatgggtttaacagccgaaaatgttgccgaacagtacggcgtatcacgtgaggcccaggacgcctttgcgtatggatcgcagatg
aaagcagcaaaggcccaagaacagggcattttcgcagctgaaatactgcctcttgaaataggggacgaagttattactcaggacgagggg
gttcgtcaagagaccaccctcgaaaaattaagtctgcttcggaccatttttaaagaagatggtactgttacagcgggcaacgcctcaacgatc
aatgatggcgcctcagccgtgatcattgcatcaaaggagtttgctgagacaaaccagattccctaccttgcgatcgtacatgatattacagag
ataggcattgatccatcaataatgggcattgctcccgtgagtgcgatcaataaactgatcgatcgtaaccaaattagcatggaagaaatcgat
ctctttgaaattaatgaggcatttgcagcatcctcggtggtagttcaaaaagagttaagcattcccgatgaaaagatcaatattggcggttccg
gtattgcactaggccatcctcttggcgccacaggagcgcgcattgtaaccaccctagcgcaccagttgaaacgtacacacggacgctatgg
tattgcctccctgtgcattggcggtggccttggcctagcaatattaatagaagtgcctcaggaagatcagccggttaaaaaattttatcaattgg
cccgtgaggaccgtctggctagacttcaggagcaagccgtgatcagcccagctacaaaacatgtactggcagaaatgacacttcctgaag
atattgccgacaatctgatcgaaaatcaaatatctgaaatggaaatccctcttggtgtggctttgaatctgagggtcaatgataagagttatacc
atcccactagcaactgaggaaccgagtgtaatcgctgcctgtaataatggtgcaaaaatggcaaaccacctgggcggttttcagtcagaatt
aaaagatggtttcctgcgtgggcaaattgtacttatgaacgtcaaagaacccgcaactatcgagcatacgatcacggcagagaaagcggca

atttttcgtgccgcagcgcagtcacatccatcgattgtgaaacgaggtgggggtctaaaagagatagtagtgcgtacgttcgatgatgatccg
acgttcctgtctattgatctgatagttgatactaaagacgcaatgggcgctaacatcattaacaccattctcgagggtgtagccggctttctgag
ggaaatccttaccgaagaaattctgttctctatttttatctaattacgcaaccgaatcaattgtgaccgccagctgtcgcataccttacgaagcact
gagtaaaaaaggtgatggtaaacgaatcgctgaaaaagtggctgctgcatctaaatttgcccagttagatccttatcgagctgcaacccaca
acaaaggtattatgaatggtattgaggccgtcgttttggcctcaggaaatgacacacgggcggtcgcggcagccgcacatgcgtatgcttc
acgcgatcagcactatcggggcttaagccagtggcaggttgcagaaggcgcgttacacggggagatcagtctaccacttgcactcggca
gcgttggcggtgcaattgaggtcttgcctaaagcgaaggcggcattcgaaatcatggggatcacagaggcgaaggagctggcagaagtc
acagctgcggtagggctggcgcaaaacctggcggcgttaagagcgcttgttagtgaaggaatacagcaaggtcacatgtcgctccaggct
cgctctcttgcattatcggtaggtgctacaggcaaggaagttgaaatcctggccgaaaaattacagggctctcgtatgaatcaggcgaacgc
tcagaccatactcgcagagatcagatcgcaaaaagttgaattgtga **(SEQ ID NO:47)**

*Sequence of Enterococcus faecium mvaE*

atgaccatgaacgttggaatcgataaaatgtcattctttgttccaccttactttgtggacatgactgatctggcagtagcacgggatgtcgatcc
caataagtttctgattggtattggccaggaccagatggcagttaatccgaaaacgcaggatattgtgacatttgccacaaatgctgccaaaaa
catactgtcagctgaggaccttgataaaattgatatggtcatagtcggcaccgagagtggaatcgatgaatccaaagcgagtgccgtagtgc
ttcacaggttgctcggtatccagaagtttgctcgctcctttgaaatcaaagaagcctgttatgggggtaccgcggctttacagttcgctgtaaac
cacattaggaatcatcctgaatcaaaggttcttgtagttgcatcagatatcgcgaaatacggcctggcttctggaggtgaaccaacgcaaggt
gcaggcgctgtggctatgctcgtctcaactgaccctaagatcattgctttcaacgacgatagcctcgcgcttacacaagatatctatgacttct
ggcgaccagttggacatgactatcctatggtcgacgggcctcttagtacagagacctacatccagtcatttcagaccgtatggcaggaatac
acaaaacggtcgcagcatgcactggcagactttgctgcccttagctttcatatcccgtatactaaaatgggcaaaaaggcgctgcttgcaatc
cttgaaggcgaatcagaggaggctcagaaccgtatactagcaaaatatgaaaagagtatagcctactccagaaaggcgggtaacctgtata
ccggtagcctgtatctaggacttatttcacttctggaaaatgcagaagaccttaaagctggtgatttaataggcctcttttcttacggttccggtg
ctgttgcggagtttttctcaggaaggctggttgaggactatcaggaacagctacttaaaacaaaacatgccgaacagctggcccatagaaag
caactgacaatcgaggagtacgaaacgatgttctccgatcgcttggacgtggacaaagacgccgaatacgaagacacattagcttatagca
tttcgtcagtccgaaacaccgtacgtgagtacaggagttga **(SEQ ID NO:48)**

*Sequence of Enterococcus gallinarum EG2 mvaE*

atgaaagaagtggttatgattgatgcggctcgcacacccattgggaaatacagaggtagtcttagtccttttacagcggtggagctggggac
actggtcacgaaagggctgctggataaaacaaagcttaagaaagacaagatagaccaagtgatattcggcaatgtgcttcaggcaggaaa
cggacaaaacgttgcaagacaaatagccctgaacagtggcttaccagttgacgtgccggcgatgactattaacgaagtttgcgggtccgga
atgaaagcggtgattttagcccgccagttaatacagttaggggaggcagagttggtcattgcaggggtacggagtcaatgtcacaagcac
ccatgctgaaaccttaccagtcagagaccaacgaatacggagagccgatatcatcaatggttaatgacgggctgacggatgcgttttccaat

gctcacatgggtcttactgccgaaaaggtggcgacccagttttcagtgtcgcgcgaggaacaagaccggtacgcattgtccagccaattga
aagcagcgcacgcggttgaagccggggtgttctcagaagagattattccggttaagattagcgacgaggatgtcttgagtgaagacgagg
cagtaagaggcaacagcactttggaaaaactgggcaccttgcggacggtgttttctgaagagggcacggttaccgctggcaatgcttcacc
gctgaatgacggcgctagtgtcgtgattcttgcatcaaaagaatacgcggaaaacaataatctgccttacctggcgacgataaaggaggttg
cggaagttggtatcgatccttctatcatgggtattgccccaataaaggccattcaaaagttaacagatcggtcgggcatgaacctgtccacga
ttgatctgttcgaaattaatgaagcattcgcggcatctagcattgttgtttctcaagagctgcaattggacgaagaaaaagtgaatatctatggc
ggggcgatagctttaggccatccaatcggcgcaagcggagcccggatactgacaaccttagcatacggcctcctgcgtgagcaaaagcg
ttatggtattgcgtcattatgtatcggcggtggtcttggtctggccgtgctgttagaagctaatatggagcagacccacaaagacgttcagaag
aaaaagttttaccagcttacccccctccgagcggagatcgcagcttatcgagaagaacgttctgactcaagaaacggcacttattttccagga
gcagacgttgtccgaagaactgtccgatcacatgattgagaatcaggtctccgaagtggaaattccaatgggaattgcacaaaattttcagat
taatggcaagaaaaaatggattcctatggcgactgaagaaccttcagtaatagcggcagcatcgaacggcgccaaaatctgcgggaacatt
tgcgcggaaacgcctcagcggcttatgcgcgggcagattgtcctgtctggcaaatcagaatatcaagccgtgataaatgccgtgaatcatc
gcaaagaagaactgattctttgcgcaaacgagtcgtacccgagtattgttaaacgcgggggaggtgttcaggatatttctacgcgggagttta
tgggttcttttcacgcgtatttatcaatcgactttctggtggacgtcaaggacgcaatgggggcaaacatgatcaactctattctcgaaagcgtt
gcaaataaactgcgtgaatggttcccggaagaggaaatactgttctccatcctgtcaaacttcgctacggagtccctggcatctgcatgttgc
gagattccttttgaaagacttggtcgtaacaaagaaattggtgaacagatcgccaagaaaattcaacaggcaggggaatatgctaagcttga
cccttaccgcgcggcaacccataacaaggggattatgaacggtatcgaagccgtcgttgccgcaacgggaaacgacacacgggctgtttc
cgcttctattcacgcatacgccgcccgtaatggcttgtaccaaggtttaacggattggcagatcaagggcgataaactggttggtaaattaac
agtcccactggctgtggcgactgtcggtggcgcgtcgaacatattaccaaaagccaaagcttccctcgccatgctggatattgattccgcaa
aagaactggcccaagtgatcgccgcggtaggtttagcacagaatctggcggcgttacgtgcattagtgacagaaggcattcagaaaggac
acatgggcttgcaagcacgttctttagcgatttcgataggtgccatcggtgaggagatagagcaagtcgcgaaaaaactgcgtgaagctga
aaaaatgaatcagcaaacggcaatacagattttagaaaaaattcgcgagaaatga **(SEQ ID NO:49)**

*Sequence of Enterococcus casseliflavus mvaE*

atgaaaatcggtattgaccgtctgtccttcttcatcccgaatttgtatttggacatgactgagctggcagaatcacgcggggatgatccagcta
aatatcatattggaatcggacaagatcagatggcagtgaatcgcgcaaacgaggacatcataacactgggtgcaaacgctgcgagtaaga
tcgtgacagagaaagaccgcgagttgattgatatggtaatcgttggcacggaatcaggaattgaccactccaaagcaagcgccgtgattatt
caccatctccttaaaattcagtcgttcgcccgttctttcgaggtaaaagaagcttgctatggcggaactgctgccctgcacatggcgaaggag
tatgtcaaaaatcatccggagcgtaaggtcttggtaattgcgtcagacatcgcgcgttatggtttggccagcggaggagaagttactcaagg
cgtgggggccgtagccatgatgattacacaaaaccccccggattctttcgattgaagacgatagtgtttttctcacagaggatatctatgatttct
ggcggcctgattactccgagttccctgtagtggacgggcccctttcaaactcaacgtatatagagagttttcagaaagtttggaaccggcaca
aggaattgtccggaagagggctggaagattatcaagctattgcttttcacataccctatacgaagatgggtaagaaagcgctccagagtgttt

tagaccaaaccgatgaagataaccaggagcgcttaatggctagatatgaggagtctattcgctatagccggagaattggtaacctgtacaca ggcagcttgtaccttggtcttacaagcttgttggaaaactctaaaagtttacaaccgggagatcggatcggcctcttttcctatggcagtggtg cggtgtccgagttctttaccgggtatttagaagaaaattaccaagagtacctgttcgctcaaagccatcaagaaatgctggatagccggactc ggattacggtcgatgaatacgagaccatcttttcagagactctgccagaacatggtgaatgcgccgaatatacgagcgacgtccccttttcta taaccaagattgagaacgacattcgttattataaaatctga **(SEQ ID NO:50)**

**[0141]** The mvaE nucleic acid can be expressed in a recombinant cell on a multicopy plasmid. The plasmid can be a high copy plasmid, a low copy plasmid, or a medium copy plasmid. Alternatively, the mvaE nucleic acid can be integrated into the host cell's chromosome. For both heterologous expression of an mvaE nucleic acid on a plasmid or as an integrated part of the host cell's chromosome, expression of the nucleic acid can be driven by either an inducible promoter or a constitutively expressing promoter. The promoter can be a strong driver of expression, it can be a weak driver of expression, or it can be a medium driver of expression of the mvaE nucleic acid.

**[0142]** The *mvaS* gene encodes a polypeptide that possesses HMG-CoA synthase activity. This polypeptide can convert acetoacetyl CoA to 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). Exemplary mvaS polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein that have at least one activity of a mvaS polypeptide.

**[0143]** Mutant mvaS polypeptides include those in which one or more amino acid residues have undergone an amino acid substitution while retaining mvaS polypeptide activity (*i.e.*, the ability to convert acetoacetyl CoA to 3-hydroxy-3-methylglutaryl-CoA). Amino acid substitutions in the mvaS polypeptide can be introduced to improve the functionality of the molecule. For example, amino acid substitutions that increase the binding affinity of the mvaS polypeptide for its substrate, or that improve its ability to convert acetoacetyl CoA to 3-hydroxy-3-methylglutaryl-CoA can be introduced into the mvaS polypeptide. In some aspects, the mutant mvaS polypeptides contain one or more conservative amino acid substitutions.

**[0144]** Standard methods, such as those described in Quant et al. (Biochem J., 1989, 262:159-164), can be used to determine whether a polypeptide has mvaS activity, by measuring HMG-CoA synthase activity. In an exemplary assay, HMG-CoA synthase activity can be

assayed by spectrophotometrically measuring the disappearance of the enol form of acetoacetyl-CoA by monitoring the change of absorbance at 303 nm. A standard 1 ml assay system containing 50 mm-Tris/HCl, pH 8.0, 10 mM-MgCl2 and 0.2 mM-dithiothreitol at 30 °C; 5 mM-acetyl phosphate, 10,M-acetoacetyl- CoA and 5 µ1 samples of extracts can be added, followed by simultaneous addition of acetyl-CoA (100 µM) and 10 units of PTA. HMG-CoA synthase activity is then measured as the difference in the rate before and after acetyl-CoA addition. The absorption coefficient of acetoacetyl-CoA under the conditions used (pH 8.0, 10 mM-$MgCl_2$), is 12.2 x $10^3$ $M^{-1}$ $cm^{-1}$. By definition, 1 unit of enzyme activity causes 1 µmol of acetoacetyl-CoA to be transformed per minute.

**[0145]** Alternatively, production of mevalonate in recombinant cells can be measured by, without limitation, gas chromatography (see U.S. Patent Application Publication No.: US 2005/0287655 A1) or HPLC (See U.S. Patent Application Publication No.: 2011/0159557 A1). As an exemplary assay, cultures can be inoculated in shake tubes containing LB broth supplemented with one or more antibiotics and incubated for 14h at 34°C at 250 rpm. Next, cultures can be diluted into well plates containing TM3 media supplemented with 1% Glucose, 0.1% yeast extract, and 200 µM IPTG to final OD of 0.2. The plate are then sealed with a Breath Easier membrane (Diversified Biotech) and incubated at 34°C in a shaker/incubator at 600 rpm for 24 hours. 1 mL of each culture is then centrifuged at 3,000 x g for 5 min. Supernatant is then added to 20% sulfuric acid and incubated on ice for 5 min. The mixture is then centrifuged for 5 min at 3000 x g and the supernatant was collected for HPLC analysis. The concentration of mevalonate in samples is determined by comparison to a standard curve of mevalonate (Sigma). The glucose concentration can additionally be measured by performing a glucose oxidase assay according to any method known in the art. Using HPLC, levels of mevalonate can be quantified by comparing the refractive index response of each sample versus a calibration curve generated by running various mevonate containing solutions of known concentration.

**[0146]** Exemplary mvaS nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of a mvaS polypeptide. Exemplary mvaS polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described

herein. Exemplary mvaS nucleic acids include, for example, mvaS nucleic acids isolated from *Listeria grayi_DSM 20601, Enterococcus faecium, Enterococcus gallinarum EG2, Enterococcus faecalis,* and/or *Enterococcus casseliflavus.* The mvaS nucleic acid encoded by the *Listeria grayi_DSM 20601 mvaS* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:51. The mvaS nucleic acid encoded by the *Enterococcus faecium mvaS* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:52. The mvaS nucleic acid encoded by the *Enterococcus gallinarum EG2 mvaS* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:53. The mvaS nucleic acid encoded by the *Enterococcus casseliflavus mvaS* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to SEQ ID NO:54. The mvaS nucleic acid encoded by the *Enterococcus faecalis mvaS* gene can have at least about 99%, 98%, 97%, 96%, 95%, 95%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, or 85% sequence identity to to the *mvaE* gene previously disclosed in *E. coli* to produce mevalonate (see US 2005/0287655 A1; Tabata, K. and Hashimoto,S.-I. Biotechnology Letters 26: 1487–1491, 2004).

*Sequence of Listeria grayi DSM 20601 mvaS*

atggaagaagtggtaattatagatgcacgtcggactccgattggtaaatatcacgggtcgttgaagaagttttcagcggtggcgctggggac
ggccgtggctaaagacatgttcgaacgcaaccagaaaatcaaagaggagatcgcgcaggtcataattggtaatgtcttgcaggcaggaaa
tggccagaaccccgcgcggcaagttgctcttcaatcagggttgtccgttgacattcccgcttctacaattaacgaggtttgtgggtctggtttg
aaagctatcttgatgggcatggaacaaatccaactcggcaaagcgcaagtagtgctggcaggcggcattgaatcaatgacaaatgcgcca
agcctgtcccactataacaaggcggaggatacgtatagtgtcccagtgtcgagcatgacactggatggtctgacagacgcattttctagtaa
acctatgggattaacagcggaaaacgtcgcacagcgctacggtatctcccgtgaggcgcaagatcaattcgcatatcaatctcagatgaaa
gcagcaaaagcgcaggcagaaaacaaattcgctaaggaaattgtgccactggcgggtgaaactaaaaccatcacagctgacgaagggat
cagatcccaaacaacgatggagaaactggcaagtctcaaacctgtttttaaaaccgatggcactgtaaccgcagggaatgctagcaccatt
aatgacggggccgcccttgtgctgcttgctagcaaaacttactgcgaaactaatgacataccgtaccttgcgacaatcaaagaaattgttgaa
gttggaatcgatccggagattatgggcatctctccgataaaagcgatacaaacattgttacaaaatcaaaaagttagcctcgaagatattgga
gtttttgaaataaatgaagcctttgccgcaagtagcatagtggttgaatctgagttgggattagatccggctaaagttaaccgttatgggggtg
gtatatccttaggtcatgcaattggggcaaccggcgctcgcctggccacttcactggtgtatcaaatgcaggagatacaagcacgttatggta
ttgcgagcctgtgcgttggtggtggacttggactggcaatgcttttagaacgtccaactattgagaaggctaaaccgacagacaaaaagttct
atgaattgtcaccagctgaacggttgcaagagctggaaaatcaacagaaaatcagttctgaaactaaacagcagttatctcagatgatgcttg

ccgaggacactgcaaaccatttgatagaaaatcaaatatcagagattgaactcccaatgggcgtcgggatgaacctgaaggttgatgggaa
agcctatgttgtgccaatggcgacggaagagccgtccgtcatcgcggccatgtctaatggtgccaaaatggccggcgaaattcacactcag
tcgaaagaacggctgctcagaggtcagattgttttcagcgcgaagaatccgaatgaaatcgaacagagaatagctgagaaccaagctttga
ttttcgaacgtgccgaacagtcctatccttccattgtgaaaagagagggaggtctccgccgcattgcacttcgtcattttcctgccgattctcag
caggagtctgcggaccagtccacatttttatcagtggacctttttgtagatgtgaaagacgcgatgggggcaaatatcataaatgcaatacttg
agggcgtcgcagccctgtttcgcgaatggttccccaatgaggaaattctttttttctattctctcgaacttggctacggagagcttagtcacggct
gtttgtgaagtcccatttagtgcacttagcaagagaggtggtgcaacggtggcccagaaaattgtgcaggcgtcgctcttcgcaaagacag
acccataccgcgcagtgacccacaacaaagggattatgaacggtgtagaggctgttatgcttgccacaggcaacgacacgcgcgcagtct
cagccgcttgtcatggatacgcagcgcgcaccggtagctatcagggtctgactaactggacgattgagtcggatcgcctggtaggcgaga
taacactgccgctggccatcgctacagttggaggcgctaccaaagtgttgcccaaagctcaagcggcactggagattagtgatgttcactct
tctcaagagcttgcagccttagcggcgtcagtaggtttagtacaaaatctcgcggccctgcgcgcactggtttccgaaggtatacaaaaagg
gcacatgtccatgcaagcccggtctctcgcaatcgcggtcggtgctgaaaaagccgagatcgagcaggtcgccgaaaagttgcggcaga
acccgccaatgaatcagcagcaggcgctccgttttcttggcgagatccgcgaacaatga **(SEQ ID NO:51)**

*Sequence of Enterococcus faecium mvaS*

atgaacgtcggcattgacaaaattaatttttcgttccaccgtattatctggatatggtcgacctggcccacgcacgcgaagtggacccgaac
aaatttacaattggaattggacaggatcagatggctgtgagcaaaaagacgcacgatatcgtaacattcgcggctagtgccgcgaaggaaa
ttttagaacctgaggacttgcaagctatagacatggttatagttggtaccgaatcgggcattgacgagagcaaagcatccgcggtcgttttac
atcgtttgttgggcgtacaacctttcgctcgcagttttgaaattaaagaagcctgttacggggcaaccgcaggcattcagtttgccaagactca
tatacaagcgaacccggagagcaaggtcctggtaattgcaagcgatatagctcggtatggtcttcggtcaggtggagagcccacacaagg
cgcaggggcagttgctatgcttctcacggcaaatcccagaatcctgaccttcgaaaacgacaatctgatgttaacgcaggatatttatgacttc
tggagaccacttggtcacgcttaccctatggtagatggccacctttccaatcaagtctatattgacagttttaagaaggtctggcaagcacattg
cgaacgcaatcaagcttctatatccgactatgccgcgattagttttcatattccgtatacaaaaatgggtaagaaagccctgctcgctgtttttgc
agatgaagtggaaactgaacaggaacgcgttatggcacggtatgaagagtctatcgtatattcacgccggatcggcaacttgtatacgggat
cattgtacctggggctgatatccttattggaaaacagttctcacctgtcggcgggcgaccggataggattgtttagttatgggagtggcgctgt
cagcgaatttttctccggtcgtttagtggcaggctatgaaaatcaattgaacaaagaggcgcatacccagctcctggatcagcgtcagaagc
tttccatcgaagagtatgaggcgatttttacagattccttagaaattgatcaggatgcagcgttctcggatgacctgccatattccatccgcgag
ataaaaaacacgattcggtactataaggagagctga **(SEQ ID NO:52)**

*Sequence of Enterococcus gallinarum EG2 mvaS*

atggaagaagttgtcatcattgacgcactgcgtactccaataggaaagtaccacggttcgctgaaagattacacagctgttgaactggggac
agtagcagcaaaggcgttgctggcacgaaatcagcaagcaaaagaacacatagcgcaagttattattggcaacgtcctgcaagccggaa
gtgggcagaatccaggccgacaagtcagtttacagtcaggattgtcttctgatatccccgctagcacgatcaatgaagtgtgtggctcgggt

atgaaagcgattctgatgggtatggagcaaattcagctgaacaaagcctctgtggtcttaacaggcggaattgaaagcatgaccaacgcgc
cgctgtttagttattacaacaaggctgaggatcaatattcggcgccggttagcacaatgatgcacgatggtctaacagatgctttcagttccaa
accaatgggcttaaccgcagagaccgtcgctgagagatatggaattacgcgtaaggaacaagatgaatttgcttatcactctcaaatgaagg
cggccaaagcccaggcggcgaaaaagtttgatcaggaaattgtacccctgacggaaaaatccggaacggttctccaggacgaaggcatc
agagccgcgacaacagtcgagaagctagctgagcttaaaacggtgttcaaaaaagacggaacagttacagcgggtaacgcctctacgat
aaatgatggcgctgctatggtattaatagcatcaaaatcttattgcgaagaacaccagattccttatctggccgttataaaggagatcgttgag
gtgggttttgcccccgaaataatgggtatttcccccattaaggctatagacaccctgctgaaaaatcaagcactgaccatagaggatatagga
atatttgagattaatgaagcctttgctgcgagttcgattgtggtagaacgcgagttgggcctggaccccaaaaaagttaatcgctatggcggt
ggtatatcactcggccacgcaattggggcgacgggagctcgcattgcgacgaccgttgcttatcagctgaaagatacccaggagcgctac
ggtatagcttccttatgcgttggtgggggtcttggattggcgatgcttctggaaaacccatcggccactgcctcacaaactaattttgatgagg
aatctgcttccgaaaaaactgagaagaagaagttttatgcgctagctcctaacgaacgcttagcgtttttggaagcccaaggcgctattaccg
ctgctgaaaccctggtcttccaggagatgaccttaaacaaagagacagccaatcacttaatcgaaaaccaaatcagcgaagttgaaattcctt
taggcgtgggcctgaacttacaggtgaatgggaaagcgtataatgttcctctggccacggaggaaccgtccgttatcgctgcgatgtcgaat
ggcgccaaaatggctggtcctattacaacaacaagtcaggagaggctgttacggggtcagattgtcttcatggacgtacaggacccagaa
gcaatattagcgaaagttgaatccgagcaagctaccattttcgcggtggcaaatgaaacatacccgtctatcgtgaaaagaggaggaggtct
gcgtagagtcattggcaggaatttcagtccggccgaaagtgacttagccacggcgtatgtatcaattgacctgatggtagatgttaaggatgc
aatgggtgctaatatcatcaatagtatcctagaaggtgttgcggaattgtttagaaaatggttcccagaagaagaaatcctgttctcaattctctc
caatctcgcgacagaaagtctggtaacggcgacgtgctcagttccgtttgataaattgtccaaaactgggaatggtcgacaagtagctggta
aaatagtgcacgcggcggactttgctaagatagatccatacagagctgccacacacaataaaggtattatgaatggcgttgaagcgttaatct
tagccaccggtaatgacacccgtgcggtgtcggctgcatgccacggttacgcggcacgcaatgggcgaatgcaagggcttacctcttgga
cgattatcgaagatcggctgataggctctatcacattacctttggctattgcgacagtgggggtgccacaaaaatcttgccaaaagcacag
gccgccctggcgctaactggcgttgagacggcgtcggaactggccagcctggcggcgagtgtgggattagttcaaaatttggccgctttac
gagcactagtgagcgagggcattcagcaagggcacatgagtatgcaagctagatccctggccattagcgtaggtgcgaaaggtactgaa
atagagcaactagctgcgaagctgagggcagcgacgcaaatgaatcaggagcaggctcgtaaatttctgaccgaaataagaaattaa
**(SEQ ID NO:53)**

*Sequence of Enterococcus casseliflavus mvaS*

atgaacgttggaattgataaaatcaatttttttcgttccgccctatttcattgatatggtggatctcgctcatgcaagagaagttgaccccaacaag
ttcactataggaataggccaagatcagatggcagtaaacaagaaaacgcaagatatcgtaacgttcgcgatgcacgccgcgaaggatattc
tgactaaggaagatttacaggccatagatatggtaatagtggggactgagtctgggatcgacgagagcaaggcaagtgctgtcgtattgcat
cggcttttaggtattcagccttttgcgcgctcctttgaaattaaggaggcatgctatggggccactgccggccttcagtttgcaaaagctcatgt
gcaggctaatccccagagcaaggtcctggtggtagcttccgatatagcacgctacggactggcatccggaggagaaccgactcaaggtgt
aggtgctgtggcaatgttgatttccgctgatccagctatcttgcagttagaaaatgataatctcatgttgacccaagatatatacgatttttggcg

cccggtcgggcatcaatatcctatggtagacggccatctgtctaatgccgtctatatagacagctttaaacaagtctggcaagcacattgcga gaaaaaccaacggactgctaaagattatgctgcattgtcgttccatattccgtacacgaaaatgggtaagaaagctctgttagcggtttttgcg gaggaagatgagacagaacaaaagcggttaatggcacgttatgaagaatcaattgtatacagtcgtcggactggaaatctgtatactggctc actctatctgggcctgatttccttactggagaatagtagcagtttacaggcgaacgatcgcataggtctgtttagctatggttcaggggccgttg cggaatttttcagtggcctcttggtaccgggttacgagaaacaattagcgcaagctgcccatcaagctcttctggacgaccggcaaaaactg actatcgcagagtacgaagccatgtttaatgaaaccattgatattgatcaggaccagtcatttgaggatgacttactgtactccatcagagaga tcaaaaacactattcgctactataacgaggagaatgaataa **(SEQ ID NO:54)**

**[0147]** The mvaS nucleic acid can be expressed in a recombinant cell on a multicopy plasmid. The plasmid can be a high copy plasmid, a low copy plasmid, or a medium copy plasmid. Alternatively, the mvaS nucleic acid can be integrated into the host cell's chromosome. For both heterologous expression of an mvaS nucleic acid on a plasmid or as an integrated part of the host cell's chromosome, expression of the nucleic acid can be driven by either an inducible promoter or a constitutively expressing promoter. The promoter can be a strong driver of expression, it can be a weak driver of expression, or it can be a medium driver of expression of the mvaS nucleic acid.

**[0148]** Compositions of recombinant cells as described herein are contemplated within the scope of the invention as well. It is understood that recombinant cells also encompass progeny cells as well.

b. Nucleic acids encoding polypeptides of the lower MVA pathway

**[0149]** In some aspects of the invention, the cells described in any of the compositions or methods described herein further comprise one or more nucleic acids encoding a lower mevalonate (MVA) pathway polypeptide(s). In some aspects, the lower MVA pathway polypeptide is an endogenous polypeptide. In some aspects, the endogenous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a constitutive promoter. In some aspects, the endogenous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to an inducible promoter. In some aspects, the endogenous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a strong promoter. In a particular aspect, the cells are engineered to over-express the endogenous lower MVA pathway polypeptide relative to wild-type cells. In some aspects, the endogenous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a weak promoter.

**[0150]** The lower mevalonate biosynthetic pathway comprises mevalonate kinase (MVK), phosphomevalonate kinase (PMK), and diphosphomevalonte decarboxylase (MVD). In some aspects, the lower MVA pathway can further comprise isopentenyl diphosphate isomerase (IDI). Cells provided herein can comprise at least one nucleic acid encoding isoprene synthase, one or more upper MVA pathway polypeptides, and/or one or more lower MVA pathway polypeptides. Polypeptides of the lower MVA pathway can be any enzyme (a) that phosphorylates mevalonate to mevalonate 5-phosphate; (b) that converts mevalonate 5-phosphate to mevalonate 5-pyrophosphate; and (c) that converts mevalonate 5-pyrophosphate to isopentenyl pyrophosphate. More particularly, the enzyme that phosphorylates mevalonate to mevalonate 5-phosphate can be from the group consisting of *M. mazei* mevalonate kinase, *Lactobacillus* mevalonate kinase polypeptide, *Methanoccoides* mevalonate kinase polypeptide, *Methanoccoides burtonii* mevalonate kinase polypeptide, *Lactobacillus sakei* mevalonate kinase polypeptide, *yeast* mevalonate kinase polypeptide, *Saccharomyces cerevisiae* mevalonate kinase polypeptide, *Streptococcus* mevalonate kinase polypeptide, *Streptococcus pneumoniae* mevalonate kinase polypeptide, *Streptomyces* mevalonate kinase polypeptide, and *Streptomyces CL190* mevalonate kinase polypeptide. In another aspect, the enzyme that phosphorylates mevalonate to mevalonate 5-phosphate is *M. mazei* mevalonate kinase.

**[0151]** In some aspects, the lower MVA pathway polypeptide is a heterologous polypeptide. In some aspects, the cells comprise more than one copy of a heterologous nucleic acid encoding a lower MVA pathway polypeptide. In some aspects, the heterologous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a constitutive promoter. In some aspects, the heterologous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to an inducible promoter. In some aspects, the heterologous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a strong promoter. In some aspects, the heterologous nucleic acid encoding a lower MVA pathway polypeptide is operably linked to a weak promoter. In some aspects, the heterologous lower MVA pathway polypeptide is a polypeptide from *Saccharomyces cerevisiae, Enterococcus faecalis,* or *Methanosarcina mazei*.

**[0152]** The nucleic acids encoding a lower MVA pathway polypeptide(s) can be integrated into a genome of the cells or can be stably expressed in the cells. The nucleic acids encoding a lower MVA pathway polypeptide(s) can additionally be on a vector.

**[0153]** Exemplary lower MVA pathway polypeptides are also provided below: (i) mevalonate kinase (MVK); (ii) phosphomevalonate kinase (PMK); (iii) diphosphomevalonate decarboxylase (MVD); and (iv) isopentenyl diphosphate isomerase (IDI). In particular, the lower MVK polypeptide can be from the genus *Methanosarcina* and, more specifically, the lower MVK polypeptide can be from *Methanosarcina mazei*. In other aspects, the lower MVK polypeptide can be from the genus *Methanococcoides* and, more specifically, from *M. burtonii*. Additional examples of lower MVA pathway polypeptides can be found in U.S. Patent Application Publication 2010/0086978 the contents of which are expressly incorporated herein by reference in their entirety with respect to lower MVK pathway polypeptides and lower MVK pathway polypeptide variants.

**[0154]** Any one of the cells described herein can comprise IDI nucleic acid(s) (*e.g.*, endogenous or heterologous nucleic acid(s) encoding IDI). Isopentenyl diphosphate isomerase polypeptides (isopentenyl-diphosphate delta-isomerase or IDI) catalyzes the interconversion of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (*e.g.*, converting IPP into DMAPP and/or converting DMAPP into IPP). Exemplary IDI polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of an IDI polypeptide. Standard methods (such as those described herein) can be used to determine whether a polypeptide has IDI polypeptide activity by measuring the ability of the polypeptide to interconvert IPP and DMAPP *in vitro*, in a cell extract, or *in vivo*. Exemplary IDI nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of an IDI polypeptide. Exemplary IDI polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein.

**[0155]** Lower MVA pathway polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of a lower MVA pathway polypeptide. Exemplary lower MVA pathway nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of a lower MVA pathway polypeptide. Exemplary lower MVA pathway polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the

source organisms described herein. In addition, variants of lower MVA pathway polypeptides that confer the result of better isoprene production can also be used as well.

**[0156]** In some aspects, the lower MVA pathway polypeptide is a polypeptide from *Saccharomyces cerevisiae, Enterococcus faecalis,* or *Methanosarcina mazei*. In some aspects, the MVK polypeptide is selected from the group consisting of *Lactobacillus* mevalonate kinase polypeptide, *Lactobacillus sakei* mevalonate kinase polypeptide, yeast mevalonate kinase polypeptide, *Saccharomyces cerevisiae* mevalonate kinase polypeptide, *Streptococcus* mevalonate kinase polypeptide, *Streptococcus pneumoniae* mevalonate kinase polypeptide, *Streptomyces* mevalonate kinase polypeptide, *Streptomyces* CL190 mevalonate kinase polypeptide, *M. burtonii* mevalonate kinase polypeptide, and *Methanosarcina mazei* mevalonate kinase polypeptide. Any one of the promoters described herein (*e.g.*, promoters described herein and identified in the Examples of the present disclosure including inducible promoters and constitutive promoters) can be used to drive expression of any of the MVA polypeptides described herein.

### 3. DXP pathway nucleic acids and polypeptides

**[0157]** In some aspects of the invention, the recombinant cells described in any of the compositions or methods described herein (including host cells that have been modified as described herein) further comprise one or more heterologous nucleic acids encoding a DXS polypeptide and/or other DXP pathway polypeptides. In some aspects, the cells further comprise a chromosomal copy of an endogenous nucleic acid encoding a DXS polypeptide and/or other DXP pathway polypeptides. In some aspects, the *E. coli* cells further comprise one or more nucleic acids encoding an IDI polypeptide and a DXS polypeptide and/or other DXP pathway polypeptides. In some aspects, one nucleic acid encodes the isoprene synthase polypeptide, IDI polypeptide, and DXS polypeptide and/or other DXP pathway polypeptides. In some aspects, one plasmid encodes the isoprene synthase polypeptide, IDI polypeptide, and DXS polypeptide and/or other DXP pathway polypeptides. In some aspects, multiple plasmids encode the isoprene synthase polypeptide, IDI polypeptide, and DXS polypeptide or other DXP pathway polypeptides.

**[0158]** Exemplary DXS polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of a DXS polypeptide. Standard

methods (such as those described herein) can be used to determine whether a polypeptide has DXS polypeptide activity by measuring the ability of the polypeptide to convert pyruvate and D-glyceraldehyde-3-phosphate into 1-deoxy-D-xylulose-5-phosphate *in vitro*, in a cell extract, or *in vivo*. Exemplary DXS polypeptides and nucleic acids and methods of measuring DXS activity are described in more detail in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (US Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, and US Publ. No. 2010/0003716.

**[0159]** Exemplary DXP pathways polypeptides include, but are not limited to any of the following polypeptides: DXS polypeptides, DXR polypeptides, MCT polypeptides, CMK polypeptides, MCS polypeptides, HDS polypeptides, HDR polypeptides, and polypeptides (*e.g.*, fusion polypeptides) having an activity of one, two, or more of the DXP pathway polypeptides. In particular, DXP pathway polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of a DXP pathway polypeptide. Exemplary DXP pathway nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of a DXP pathway polypeptide. Exemplary DXP pathway polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein. Exemplary DXP pathway polypeptides and nucleic acids and methods of measuring DXP pathway polypeptide activity are described in more detail in International Publication No.: WO 2010/148150

**[0160]** Exemplary DXS polypeptides include polypeptides, fragments of polypeptides, peptides, and fusions polypeptides that have at least one activity of a DXS polypeptide. Standard methods (such as those described herein) can be used to determine whether a polypeptide has DXS polypeptide activity by measuring the ability of the polypeptide to convert pyruvate and D-glyceraldehyde-3-phosphate into 1-deoxy-D-xylulose-5-phosphate *in vitro*, in a cell extract, or *in vivo*. Exemplary DXS polypeptides and nucleic acids and methods of measuring DXS activity are described in more detail in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (US Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, and US Publ. No. 2010/0003716.

**[0161]** In particular, DXS polypeptides convert pyruvate and D-glyceraldehyde 3-phosphate into 1-deoxy-d-xylulose 5-phosphate (DXP). Standard methods can be used to determine whether a polypeptide has DXS polypeptide activity by measuring the ability of the polypeptide to convert pyruvate and D-glyceraldehyde 3-phosphate *in vitro*, in a cell extract, or *in vivo*.

**[0162]** DXR polypeptides convert 1-deoxy-d-xylulose 5-phosphate (DXP) into 2-C-methyl-D-erythritol 4-phosphate (MEP). Standard methods can be used to determine whether a polypeptide has DXR polypeptides activity by measuring the ability of the polypeptide to convert DXP *in vitro*, in a cell extract, or *in vivo*.

**[0163]** MCT polypeptides convert 2-C-methyl-D-erythritol 4-phosphate (MEP) into 4-(cytidine 5'-diphospho)-2-methyl-D-erythritol (CDP-ME). Standard methods can be used to determine whether a polypeptide has MCT polypeptides activity by measuring the ability of the polypeptide to convert MEP *in vitro*, in a cell extract, or *in vivo*.

**[0164]** CMK polypeptides convert 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME) into 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP). Standard methods can be used to determine whether a polypeptide has CMK polypeptides activity by measuring the ability of the polypeptide to convert CDP-ME *in vitro*, in a cell extract, or *in vivo*.

**[0165]** MCS polypeptides convert 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP) into 2-C-methyl-D-erythritol 2, 4-cyclodiphosphate (ME-CPP or cMEPP). Standard methods can be used to determine whether a polypeptide has MCS polypeptides activity by measuring the ability of the polypeptide to convert CDP-MEP *in vitro*, in a cell extract, or *in vivo*.

**[0166]** HDS polypeptides convert 2-C-methyl-D-erythritol 2, 4-cyclodiphosphate into (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate (HMBPP or HDMAPP). Standard methods can be used to determine whether a polypeptide has HDS polypeptides activity by measuring the ability of the polypeptide to convert ME-CPP *in vitro*, in a cell extract, or *in vivo*.

**[0167]** HDR polypeptides convert (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In one embodiment, the ispH gene can be used to encode for HDR polypeptides. IspH is also known as 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase, 4Fe-4S protein, ECK0030, JW0027, lytB, yaaE,

and b0029. Standard methods can be used to determine whether a polypeptide has HDR polypeptides activity by measuring the ability of the polypeptide to convert HMBPP *in vitro*, in a cell extract, or *in vivo*.

## 4. Source organisms for MVA pathway, isoprene synthase, IDI, and DXP pathway polypeptides

**[0168]** Isoprene synthase, IDI, DXP pathway, and/or MVA pathway nucleic acids can be obtained from any organism that naturally contains isoprene synthase, IDI, DXP pathway, and/or MVA pathway nucleic acids. Isoprene is formed naturally by a variety of organisms, such as bacteria, yeast, plants, and animals. Some organisms contain the MVA pathway for producing isoprene. Isoprene synthase nucleic acids can be obtained, *e.g.*, from any organism that contains an isoprene synthase. MVA pathway nucleic acids can be obtained, *e.g.*, from any organism that contains the MVA pathway. IDI and DXP pathway nucleic acids can be obtained, *e.g.*, from any organism that contains the IDI and DXP pathway.

**[0169]** The nucleic acid sequence of the isoprene synthase, DXP pathway, IDI, and/or MVA pathway nucleic acids can be isolated from a bacterium, fungus, plant, algae, or cyanobacterium. Exemplary source organisms include, for example, yeasts, such as species of *Saccharomyces* (*e.g.*, *S. cerevisiae*), bacteria, such as species of *Escherichia* (*e.g.*, *E. coli*), or species of *Methanosarcina* (*e.g.*, *Methanosarcina mazei*), plants, such as kudzu or poplar (*e.g.*, *Populus alba* or *Populus alba x tremula* CAC35696) or aspen (*e.g.*, *Populus tremuloides*). Exemplary sources for isoprene synthases, IDI, and/or MVA pathway polypeptides which can be used are also described in International Patent Application Publication Nos. WO 2009/076676, WO 2010/003007, WO 2009/132220, WO 2010/031062, WO 2010/031068, WO 2010/031076, WO 2010/013077, WO 2010/031079, WO 2010/148150, WO 2010/078457, and WO 2010/148256.

**[0170]** In some aspects, the source organism is a yeast, such as *Saccharomyces sp.*, *Schizosaccharomyces sp.*, *Pichia sp.*, or *Candida sp.*

**[0171]** In some aspects, the source organism is a bacterium, such as strains of *Bacillus* such as *B. lichenformis* or *B.* subtilis, strains of *Pantoea* such as *P. citrea*, strains of *Pseudomonas* such as *P. alcaligenes,* strains of *Streptomyces* such as *S. lividans* or *S. rubiginosus,* strains of

*Escherichia* such as *E. coli,* strains of *Enterobacter*, strains of *Streptococcus*, or strains of *Archaea* such as *Methanosarcina mazei*.

**[0172]** As used herein, "the genus *Bacillus*" includes all species within the genus "*Bacillus*," as known to those of skill in the art, including but not limited to *B. subtilis*, *B. licheniformis, B. lentus, B. brevis, B.* stearothermophilus, *B. alkalophilus, B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterium, B. coagulans, B. circulans, B. lautus*, and *B. thuringiensis*. It is recognized that the genus *Bacillus* continues to undergo taxonomical reorganization. Thus, it is intended that the genus include species that have been reclassified, including but not limited to such organisms as *B. stearothermophilus*, which is now named "*Geobacillus stearothermophilus*." The production of resistant endospores in the presence of oxygen is considered the defining feature of the genus *Bacillus*, although this characteristic also applies to the recently named *Alicyclobacillus*, *Amphibacillus*, *Aneurinibacillus*, *Anoxybacillus*, *Brevibacillus*, *Filobacillus*, *Gracilibacillus*, *Halobacillus*, *Paenibacillus*, *Salibacillus*, *Thermobacillus*, *Ureibacillus*, and *Virgibacillus*.

**[0173]** In some aspects, the source organism is a gram-positive bacterium. Non-limiting examples include strains of Streptomyces (*e.g.*, *S. lividans*, *S. coelicolor*, or *S. griseus*) and *Bacillus*. In some aspects, the source organism is a gram–negative bacterium, such as *E. coli* or *Pseudomonas sp.* In some aspects, the source organism is *L. acidophilus*.

**[0174]** In some aspects, the source organism is a plant, such as a plant from the family Fabaceae, such as the Faboideae subfamily. In some aspects, the source organism is kudzu, poplar (such as *Populus alba x tremula* CAC35696), aspen (such as *Populus tremuloides*), or *Quercus robur*.

**[0175]** In some aspects, the source organism is an algae, such as a green algae, red algae, glaucophytes, chlorarachniophytes, euglenids, chromista, or dinoflagellates.

**[0176]** In some aspects, the source organism is a cyanobacteria, such as cyanobacteria classified into any of the following groups based on morphology: *Chroococcales*, *Pleurocapsales*, *Oscillatoriales*, *Nostocales*, or *Stigonematales*.

## 5. Phosphoketolase nucleic acids and polypeptides

**[0177]** In some aspects of the invention, the recombinant cells described in any of the compositions or methods described herein further comprise one or more nucleic acids encoding a phosphoketolase polypeptide or a polypeptide having phosphoketolase activity. In some aspects, the phosphoketolase polypeptide is an endogenous polypeptide. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a constitutive promoter. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to an inducible promoter. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a strong promoter. In some aspects, more than one endogenous nucleic acid encoding a phosphoketolase polypeptide is used (e.g, 2, 3, 4, or more copies of an endogenous nucleic acid encoding a phosphoketolase polypeptide). In a particular aspect, the cells are engineered to overexpress the endogenous phosphoketolase polypeptide relative to wild-type cells. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a weak promoter.

**[0178]** Phosphoketolase enzymes catalyze the conversion of xylulose 5-phosphate to glyceraldehyde 3-phosphate and acetyl phosphate and/or the conversion of fructose 6-phosphate to erythrose 4-phosphate and acetyl phosphate. In certain embodiments, the phosphoketolase polypeptide catalyzes the conversion of sedoheptulose-7-phosphate to a product (e.g., ribose-5-phosphate) and acetyl phosphate. Thus, without being bound by theory, the expression of phosphoketolase as set forth herein can result in an increase in the amount of acetyl phosphate produced from a carbon source. This acetyl phosphate can be converted into acetyl-CoA which can then be utilized by the enzymatic activities of the MVA pathway to produce mevalonate, and/or isoprene. Thus the amount of these compounds produced from a carbon source may be increased. In certain embodiments, the phosphoketolase enzyme is capable of catalyzing the conversion of xylulose 5-phosphate to glyceraldehyde 3-phosphate and acetyl phosphate. In other embodiments, the phosphoketolase enzyme is capable of catalyzing the conversion of fructose 6-phosphate to erythrose 4-phosphate and acetyl phosphate. Thus, without being bound by theory, the expression of phosphoketolase as set forth herein can result in an increase in the amount of acetyl phosphate produced from a carbohydrate source. This acetyl phosphate can be converted into acetyl-CoA which can then be utilized by the enzymatic activities of the MVA

pathway to produce isoprene. Thus the amount of these compounds produced from a carbohydrate substrate may be increased. Alternatively, production of Acetyl-P and AcCoA can be increased without the increase being reflected in higher intracellular concentration. In certain embodiments, intracellular acetyl-P or acetyl-CoA concentrations will remain unchanged or even decrease, even though the phosphoketolase reaction is taking place.

**[0179]** Accordingly, in certain embodiments, the recombinant cells described herein in any of the methods described herein further comprise one or more nucleic acids encoding a phosphosphoketolase polypeptide or a polypeptide having phosphoketolase activity. In some aspects, the phosphoketolase polypeptide is an endogenous polypeptide. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a constitutive promoter. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to an inducible promoter. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a strong promoter. In some aspects, more than one endogenous nucleic acid encoding a phosphoketolase polypeptide is used (e.g, 2, 3, 4, or more copies of an endogenous nucleic acid encoding a phosphoketolase polypeptide). In a particular aspect, the cells are engineered to overexpress the endogenous phosphoketolase polypeptide relative to wild-type cells. In some aspects, the endogenous nucleic acid encoding a phosphoketolase polypeptide is operably linked to a weak promoter.

**[0180]** Exemplary phosphoketolase nucleic acids include nucleic acids that encode a polypeptide, fragment of a polypeptide, peptide, or fusion polypeptide that has at least one activity of a phosphoketolase polypeptide. Exemplary phosphoketolase polypeptides and nucleic acids include naturally-occurring polypeptides and nucleic acids from any of the source organisms described herein as well as mutant polypeptides and nucleic acids derived from any of the source organisms described herein. In some aspects, a nucleic acid encoding a phosphoketolase is from *Clostridium acetobutylicum, Lactobacillus reuteri, Lactobacillus plantarum, Lactobacillus paraplantarum, Bifidobacterium longum, Bifidobacterium animalis, Bifidobacterium breve, Enterococcus gallinarum, Gardnerella vaginalis, Ferrimonas balearica, Mucilaginibacter paludis, Nostoc punctiforme, Nostoc punctiforme PCC 73102, Pantoea, Pedobactor saltans, Rahnella aquatilis, Rhodopseudomonas palustris, Streptomyces griseus, Streptomyces avermitilis, Nocardiopsis dassonvillei, and/or Thermobifida fusca.* Additional

examples of phosphoketolase enzymes which can be used herein are described in U.S. 7,785,858 and International Patent Application Publication No. WO 2011/159853 which are incorporated by reference herein.

**[0181]** Standard methods can be used to determine whether a polypeptide has phosphoketolase peptide activity by measuring the ability of the peptide to convert D-fructose 6-phosphate or D-xylulose 5-phosphate into acetyl-P. Acetyl-P can then be converted into ferryl acetyl hydroxamate, which can be detected spectrophotometrically (Meile et al., J. Bact. 183:2929-2936, 2001). Any polypeptide identified as having phosphoketolase peptide activity as described herein is suitable for use in the present invention.

**[0182]** In other aspects, exemplary phosphoketolase nucleic acids include, for example, a phosphoketolase isolated from Lactobacillus reuteri, Bifidobacterium longum, Ferrimonas balearica, Pedobactor saltans, Streptomyces griseus, and/or Nocardiopsis dassonvillei. Additional examples of phosphoketolase enzymes which can be used herein are described in U.S. 7,785,858, which is incorporated by reference herein.

**[0183]** In any of the embodiments herein, the recombinant cells can be further engineered to increase the activity of one or more of the following genes selected from the group consisting of ribose-5-phosphate isomerase (*rpiA* and/or *rpiB*), D-ribulose-5-phosphate 3-epimerase (*rpe*), transketolase (*tktA* and/or *tktB*), transaldolase B (*tal B*), phosphate acetyltransferase (*pta* and/or *eutD*). In another embodiment, the recombinant cells can be further engineered to decrease the activity of one or more genes of the following genes including glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphofructokinase-1 (*pfkA* and/or *pfkB*), fructose bisphosphate aldolase (*fba*, *fbaA*, *fbaB*, and/or *fbaC*), glyceraldehyde-3-phosphate dehydrogenase (*gapA* and/or *gapB*), acetate kinase (*ackA*), citrate synthase (*gltA*), EI (*ptsI*), $EIICB^{Glc}$ (*ptsG*), $EIIA^{Glc}$ (*crr*), and/or HPr (*ptsH*).

**Pathways involving glyceraldehyde 3-phosphate**

**[0184]** Glyceraldehyde 3-phosphate dehydrogenase (*gapA* and/or *gapB*) is a crucial enzyme of glycolysis catalyzes the conversion of glyceraldehyde 3-phosphate into 1,3-biphospho-D-glycerate (Branlant G. and Branlant C. 1985. Eur. J. Biochem. 150:61-66).

**[0185]** In certains aspects, recombinant cells comprising one or more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein further compre one more nucleic acids encoding a phosphoketolase polypeptide. In order to direct carbon towards the phosphoketolase enzyme, glyceraldehyde 3-phosphate dehydrogenase expression can be modulated (e.g., decrease enzyme activity) to allow more carbon to flux towards fructose 6-phosphate and xylulose 5-phosphate, thereby increasing the eventual production of isoprene. Decrease of glyceraldehyde 3-phosphate dehydrogenase activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%. Or 100%. In some aspects, the activity of glyceraldehyde 3-phosphate dehydrogenase is modulated by decreasing the activity of an endogenous glyceraldehyde 3-phosphate dehydrogenase. This can be accomplished by replacing the endogenous glyceraldehyde 3-phosphate dehydrogenase gene promoter with a synthetic constitutively low expressing promoter. The gene encoding glyceraldehyde 3-phosphate dehydrogenase can also be deleted. The gene encoding glyceraldehyde 3-phosphate dehydrogenase can also be replaced by a *Bacillus* enzyme catalyzing the same reaction but producing NADPH rather than NADH. The decrease of the activity of glyceraldehyde 3-phosphate dehydrogenase can result in more carbon flux into the mevalonate-dependent biosynthetic pathway in comparison to cells that do not have decreased expression of glyceraldehyde 3-phosphate dehydrogenase. In any aspects of the invention, provided herein are recombinant cells comprising one or more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to decrease the activity of glyceraldehyde 3-phosphate dehydrogenase (*gapA* and/or *gapB*). Activity modulation (e.g., decreased) of glyceraldehyde 3-phosphate dehydrogenase isozymes is also contemplated herein. In any aspects of the invention, provided herein are recom binant cells comprising one or more heterologously expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to decrease the activity of a glyceraldehyde 3-phosphate dehydrogenase (*gapA* and/or *gapB*) isozyme.

a. Pathways involving the Entner-Doudoroff pathway

**[0186]** The Entner-Doudoroff (ED) pathway is an alternative to the Emden-Meyerhoff-Parnass (EMP –glycolysis) pathway. Some organisms, like *E. coli*, harbor both the ED and EMP pathways, while others have only one or the other. *Bacillus* subtilis has only the EMP pathway, while *Zymomonas mobilis* has only the ED pathway (Peekhaus and Conway. 1998. J. Bact. 180:3495-3502; Stulke and Hillen. 2000. Annu. Rev. Microbiol. 54, 849–880; Dawes et al. 1966. Biochem. J. 98:795-803). Fructose bisphophate aldolase (fba, fbaA, fbaB, and/or fbaC) interacts with the Entner-Doudoroff pathway and reversibly catalyzes the conversion of fructose 1,6-bisphosphate into dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) (Baldwin S.A., et. al., Biochem J. (1978) 169(3):633-41).

**[0187]** Phosphogluconate dehydratase (edd) removes one molecule of $H_2O$ from 6-phospho-D-gluconate to form 2-dehydro-3-deoxy-D-gluconate 6-phosphate, while 2-keto-3-deoxygluconate 6-phosphate aldolase (eda) catalyzes an aldol cleavage (Egan et al. 1992. J. Bact. 174:4638-4646). The two genes are in an operon.

**[0188]** Metabolites that can be directed into the phosphoketolase pathway can also be diverted into the ED pathway. To avoid metabolite loss to the ED-pathway, phosphogluconate dehydratase gene (*e.g.*, the endogenous phosphogluconate dehydratase gene) and/or an 2-keto-3-deoxygluconate 6-phosphate aldolase gene (*e.g.*, the endogenous 2-keto-3-deoxygluconate 6-phosphate aldolase gene) activity is attenuated. One way of achieving attenuation is by deleting phosphogluconate dehydratase (edd) and/or 2-keto-3-deoxygluconate 6-phosphate aldolase (eda). This can be accomplished by replacing one or both genes with a chloramphenicol or kanamycin cassette followed by looping out of the cassette. Without these enzymatic activities, more carbon can flux through the phosphoketolase enzyme, thus increasing the yield of isoprene.

**[0189]** The activity of phosphogluconate dehydratase (edd) and/or 2-keto-3-deoxygluconate 6-phosphate aldolase (eda) can also be decreased by other molecular manipulations of the enzymes. The decrease of enzyme activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

**[0190]** In some cases, attenuating the activity of the endogenous phosphogluconate dehydratase gene and/or the endogenous 2-keto-3-deoxygluconate 6-phosphate aldolase gene results in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to cells that do not have attenuated endogenous phosphogluconate dehydratase gene and/or endogenous acetate kinase2-keto-3-deoxygluconate 6-phosphate aldolase gene expression.

b. Pathways involving the oxidative branch of the pentose phosphate pathway

**[0191]** *E. coli* uses the pentose phosphate pathway to break down hexoses and pentoses and to provide cells with intermediates for various anabolic pathways. It is also a major producer of NADPH. The pentose phosphate pathway is composed from an oxidative branch (with enzymes like glucose 6-phosphate 1-dehydrogenase (*zwf*), 6-phosphogluconolactonase (*pgl*) or 6-phosphogluconate dehydrogenase (*gnd*)) and a non-oxidative branch (with enzymes such as transketolase (*tktA*), transaldolase (*talA* or talB), ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase) (Sprenger. 1995. Arch. Microbiol.164:324-330).

**[0192]** In order to direct carbon towards the phosphoketolase enzyme, the non-oxidative branch of the pentose phosphate pathway (transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase) expression can be modulated (*e.g.*, increase enzyme activity) to allow more carbon to flux towards fructose 6-phosphate and xylulose 5-phosphate, thereby increasing the eventual production of isoprene. Increase of transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase activity can be any amount of increase of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the enzyme activity is increased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%. In some aspects, the activity of transketolase, transaldolase , ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase is modulated by increasing the activity of an endogenous transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase. This can be accomplished by replacing the endogenous transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase gene promoter with a synthetic constitutively high expressing promoter. The genes encoding transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate

epimerase can also be cloned on a plasmid behind an appropriate promoter. The increase of the activity of transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase can result in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to cells that do not have increased expression of transketolase, transaldolase , ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase.

c. Pathways involving phosphofructokinase

**[0193]** Phosphofructokinase is a crucial enzyme of glycolysis which catalyzes the phosphorylation of fructose 6-phosphate. E. coli has two isozymes encoded by pfkA and pfkB. Most of the phosphofructokinase activity in the cell is due to pfkA (Kotlarz et al. 1975 Biochim. Biophys. Acta 381:257-268).

**[0194]** In order to direct carbon towards the phosphoketolase enzyme, phosphofructokinase expression can be modulated (*e.g.*, decrease enzyme activity) to allow more carbon to flux towards fructose 6-phosphate and xylulose 5-phosphate, thereby increasing the eventual production of isoprene. Decrease of phosphofructokinase activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%. Or 100%. In some aspects, the activity of phosphofructokinase is modulated by decreasing the activity of an endogenous phosphofructokinase. This can be accomplished by replacing the endogenous phosphofructokinase gene promoter with a synthetic constitutively low expressing promoter. The gene encoding phosphofructokinase can also be deleted. The decrease of the activity of phosphofructokinase can result in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to cells that do not have decreased expression of phosphofructokinase.

## 6. Additional Host cell Mutations

**[0195]** The invention also contemplates additional host cell mutations that increase carbon flux through the MVA pathway. By increasing the carbon flow, more isoprene can be produced. The recombinant cells comprising acetoacetyl-CoA synthase as described herein can also be engineered for increased carbon flux towards mevalonate production wherein the activity of one

or more enzymes from the group consisting of: (a) citrate synthase, (b) phosphotransacetylase; (c) acetate kinase; (d) lactate dehydrogenase; (e) NADP-dependent malic enzyme, and; (f) pyruvate dehydrogenase; (g) 6-phosphogluconolactonase; (h) phosphoenolpyruvate carboxylase; (i) the inhibitor of RssB activity during magnesium starvation protein; (j) the acrA component of the multidrug efflux pump acrAB-TolC; and (k) the fumarate and nitrate reduction sRNA (FNR) is modulated.

a. Citrate synthase pathway

**[0196]** Citrate synthase catalyzes the condensation of oxaloacetate and acetyl-CoA to form citrate, a metabolite of the Tricarboxylic acid (TCA) cycle (Ner, S. et al. 1983. *Biochemistry*, 22: 5243-5249; Bhayana, V. and Duckworth, H. 1984. *Biochemistry* 23: 2900-2905). In *E. coli*, this enzyme, encoded by *gltA*, behaves like a trimer of dimeric subunits. The hexameric form allows the enzyme to be allosterically regulated by NADH. This enzyme has been widely studied (Wiegand, G., and Remington, S. 1986. *Annual Rev. Biophysics Biophys. Chem.*15: 97-117; Duckworth et al. 1987. *Biochem Soc Symp*. 54:83-92; Stockell, D. et al. 2003. J. Biol. Chem. 278: 35435-43; Maurus, R. et al. 2003. *Biochemistry*. 42:5555-5565). To avoid allosteric inhibition by NADH, replacement by or supplementation with the *Bacillus subtilis* NADH-insensitive citrate synthase has been considered (Underwood et al. 2002. *Appl. Environ. Microbiol.* 68:1071-1081; Sanchez et al. 2005. *Met. Eng*. 7:229-239).

**[0197]** The reaction catalyzed by citrate synthase is directly competing with the thiolase catalyzing the first step of the mevalonate pathway, as they both have acetyl-CoA as a substrate (Hedl et al. 2002. *J. Bact.* 184:2116-2122). Therefore, one of skill in the art can modulate citrate synthase expression (*e.g.*, decrease enzyme activity) to allow more carbon to flux into the mevalonate pathway, thereby increasing the eventual production of isoprene. Decrease of citrate synthase activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. In some aspects, the activity of citrate synthase is modulated by decreasing the activity of an endogenous citrate synthase gene. This can be accomplished by chromosomal replacement of an endogenous citrate synthase gene with a transgene encoding an NADH-insensitive citrate synthase or by using a transgene encoding an

NADH-insensitive citrate synthase that is derived from *Bacillus subtilis*. The activity of citrate synthase can also be modulated (*e.g.*, decreased) by replacing the endogenous citrate synthase gene promoter with a synthetic constitutively low expressing promoter. The decrease of the activity of citrate synthase can result in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to microorganisms that do not have decreased expression of citrate synthase. In any aspects of the invention, provided herein are recombinant cells comprising one or more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to decrease the activity of citrate synthase (gltA). Activity modulation (e.g., decreased) of citrate synthase isozymes is also contemplated herein. In any aspects of the invention, provided herein are recombinant cells comprising one or more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to decrease the activity of a citrate synthase isozyme.

b. Pathways involving phosphotransacetylase and/or acetate kinase

**[0198]** Phosphotransacetylase ((encoded in *E. coli* by (i) *pta* (Shimizu et al. 1969. Biochim. Biophys. Acta 191: 550-558) or (ii) *eutD* (Bologna et al. 2010. J of Microbiology. 48:629-636)) catalyzes the reversible conversion between acetyl-CoA and acetylphosphate (acetyl-P), while acetate kinase (encoded in *E. coli* by ackA) (Kakuda, H. et al. 1994. J. Biochem. 11:916-922) uses acetyl-P to form acetate. These genes can be transcribed as an operon in *E. coli*. Together, they catalyze the dissimilation of acetate, with the release of ATP. Thus, one of skill in the art can increase the amount of available acetyl Co-A by attenuating the activity of phosphotransacetylase gene (*e.g.*, the endogenous phosphotransacetylase gene) and/or an acetate kinase gene (*e.g.*, the endogenous acetate kinase gene). In certain embodiments, enhancement is achieved by placing an upregulated promoter upstream of the gene in the chromosome, or to place a copy of the gene behind an adequate promoter on a plasmid. In order to decrease the amount of acetyl-coA going towards acetate, the activity of acetate kinase gene (e.g., the endogenous acetate kinase gene) can be decreased or attenuated. One way of achieving attenuation is by deleting phosphotransacetylase (pta) and/or acetate kinase (ackA). This can be accomplished by replacing one or both genes with a chloramphenicol cassette followed by looping out of the cassette. Acetate is produced by *E. coli* for a variety of reasons (Wolfe, A. 2005. Microb. Mol. Biol. Rev. 69:12-50). Without being bound by theory, since ackA-pta use

acetyl-CoA, deleting those genes might allow carbon not to be diverted into acetate and to increase the yield of isoprene. Without being bound by theory, deletion of ackA could result in decreased carbon being diverted into acetate production (since ackA use acetyl-CoA) and thereby increase the yield of isoprene.

**[0199]** In some aspects, the recombinant microorganism produces decreased amounts of acetate in comparison to microorganisms that do not have attenuated endogenous phosphotransacetylase gene and/or endogenous acetate kinase gene expression. Decrease in the amount of acetate produced can be measured by routine assays known to one of skill in the art. The amount of acetate reduction is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% as compared when no molecular manipulations are done.

**[0200]** The activity of phosphotransacetylase (pta) and/or acetate kinase (ackA) can also be decreased by other molecular manipulation of the enzymes. The decrease of enzyme activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

**[0201]** The activity of phosphotransacetylase (pta and/or eutD) can be increased by other molecular manipulations of the enzymes. The increase of enzyme activity can be and increase in any amount of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the increase of enzyme activity is increased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. In one embodiment the activity of pta is increased by altering the promoter and/or rbs on the chromosome, or by expressing it from a plasmid. In any aspects of the invention, provided herein are recombinant cells comprising one or more heterologously expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to increase the activity of phosphotransacetylase (pta and/or eutD). Activity modulation (e.g., increased) of phosphotransacetylase isozymes is also contemplated herein. In any aspects of the invention, provided herein are recombinant cells comprising one or

more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to increase the activity of a phosphotransacetylase (pta and/or eutD) isozyme.

**[0202]** The activity of acetate kinase (ackA) can also be decreased by other molecular manipulations of the enzymes. The decrease of enzyme activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. In any aspects of the invention, provided herein are recombinant cells comprising one or more heterologously expressed nucleic acids encoding phosphoketolase polypeptides as disclosed herein and further engineered to decrease the activity of acetate kinase (ackA). Activity modulation (e.g., decreased) of acetate kinase isozymes is also contemplated herein. In any aspects of the invention, provided herein are recombinant cells comprising one or more expressed nucleic acids encoding monophosphate decarboxylase and/or isopentenyl kinase polypeptides as disclosed herein and further engineered to decrease the activity of a acetate kinase isozyme.

**[0203]** In some cases, attenuating the activity of the endogenous phosphotransacetylase gene and/or the endogenous acetate kinase gene results in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to microorganisms that do not have attenuated endogenous phosphotransacetylase gene and/or endogenous acetate kinase gene expression.

c. Pathways involving lactate dehydrogenase

**[0204]** In *E. coli*, D-Lactate is produced from pyruvate through the enzyme lactate dehydrogenase (encoded by *ldhA*) (Bunch, P. et al. 1997. Microbiol. 143:187-195). Production of lactate is accompanied with oxidation of NADH, hence lactate is produced when oxygen is limited and cannot accommodate all the reducing equivalents. Thus, production of lactate could be a source for carbon consumption. As such, to improve carbon flow through to isoprene production, one of skill in the art can modulate the activity of lactate dehydrogenase, such as by decreasing the activity of the enzyme.

**[0205]** Accordingly, in one aspect, the activity of lactate dehydrogenase can be modulated by attenuating the activity of an endogenous lactate dehydrogenase gene. Such attenuation can be achieved by deletion of the endogenous lactate dehydrogenase gene. Other ways of attenuating the activity of lactate dehydrogenase gene known to one of skill in the art may also be used. By manipulating the pathway that involves lactate dehydrogenase, the recombinant microorganism produces decreased amounts of lactate in comparison to microorganisms that do not have attenuated endogenous lactate dehydrogenase gene expression. Decrease in the amount of lactate produced can be measured by routine assays known to one of skill in the art. The amount of lactate reduction is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% as compared when no molecular manipulations are done.

**[0206]** The activity of lactate dehydrogenase can also be decreased by other molecular manipulations of the enzyme. The decrease of enzyme activity can be any amount of reduction of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the decrease of enzyme activity is decreased by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

**[0207]** Accordingly, in some cases, attenuation of the activity of the endogenous lactate dehydrogenase gene results in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to microorganisms that do not have attenuated endogenous lactate dehydrogenase gene expression.

d. Pathways involving malic enzyme

**[0208]** Malic enzyme (in *E. coli sfcA* and *maeB*) is an anaplerotic enzyme that catalyzes the conversion of malate into pyruvate (using NAD+ or NADP+) by the equation below:

$$\text{(S)-malate} + \text{NAD(P)}^+ \rightleftharpoons \text{pyruvate} + CO_2 + \text{NAD(P)H}$$

**[0209]** Thus, the two substrates of this enzyme are (S)-malate and $NAD(P)^+$, whereas its 3 products are pyruvate, $CO_2$, and NADPH.

**[0210]** Expression of the NADP-dependent malic enzyme (maeB) (Iwikura, M. et al. 1979. J. Biochem. 85: 1355-1365) can help increase isoprene yield by 1) bringing carbon from the TCA cycle back to pyruvate, direct precursor of acetyl-CoA, itself direct precursor of the mevalonate pathway and 2) producing extra NADPH which could be used in the HMG-CoA reductase reaction (Oh, MK et al. (2002) *J. Biol. Chem.* 277: 13175-13183; Bologna, F. et al. (2007) J. Bact. 189:5937-5946).

**[0211]** As such, more starting substrate (pyruvate or acetyl-CoA) for the downstream production of isoprene can be achieved by modulating, such as increasing, the activity and/or expression of malic enzyme. The NADP-dependent malic enzyme gene can be an endogenous gene. One non-limiting way to accomplish this is by replacing the endogenous NADP-dependent malic enzyme gene promoter with a synthetic constitutively expressing promoter. Another non-limiting way to increase enzyme activity is by using one or more heterologous nucleic acids encoding an NADP-dependent malic enzyme polypeptide. One of skill in the art can monitor the expression of maeB RNA during fermentation or culturing using readily available molecular biology techniques.

**[0212]** Accordingly, in some embodiments, the recombinant microorganism produces increased amounts of pyruvate in comparison to microorganisms that do not have increased expression of an NADP-dependent malic enzyme gene. In some aspects, increasing the activity of an NADP-dependent malic enzyme gene results in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to microorganisms that do not have increased NADP-dependent malic enzyme gene expression.

**[0213]** Increase in the amount of pyruvate produced can be measured by routine assays known to one of skill in the art. The amount of pyruvate increase can be at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% as compared when no molecular manipulations are done.

**[0214]** The activity of malic enzyme can also be increased by other molecular manipulations of the enzyme. The increase of enzyme activity can be any amount of increase of specific activity or total activity as compared to when no manipulation has been effectuated. In some instances, the increase of enzyme activity is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%,

9%, 10%, 15%, 20%, 25%, 30%, 35%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%.

e. Pathways involving pyruvate dehydrogenase complex

**[0215]** The pyruvate dehydrogenase complex, which catalyzes the decarboxylation of pyruvate into acetyl-CoA, is composed of the proteins encoded by the genes aceE, aceF and lpdA. Transcription of those genes is regulated by several regulators. Thus, one of skill in the art can increase acetyl-CoA by modulating the activity of the pyruvate dehydrogenase complex. Modulation can be to increase the activity and/or expression (*e.g.*, constant expression) of the pyruvate dehydrogenase complex. This can be accomplished by different ways, for example, by placing a strong constitutive promoter, like PL.6 (aattcatataaaaaacatacagataaccatctgcggtgataaattatctctggcggtgttgacataaataccactggcggtgatactgagcac atcagcaggacgcactgaccaccatgaaggtg (SEQ ID NO:30) lambda promoter, GenBank NC_001416), in front of the operon or using one or more synthetic constitutively expressing promoters.

**[0216]** Accordingly, in one aspect, the activity of pyruvate dehydrogenase is modulated by increasing the activity of one or more genes of the pyruvate dehydrogenase complex consisting of (a) pyruvate dehydrogenase (E1), (b) dihydrolipoyl transacetylase, and (c) dihydrolipoyl dehydrogenase. It is understood that any one, two or three of these genes can be manipulated for increasing activity of pyruvate dehydrogenase. In another aspect, the activity of the pyruvate dehydrogenase complex can be modulated by attenuating the activity of an endogenous pyruvate dehydrogenase complex repressor gene, further detailed below. The activity of an endogenous pyruvate dehydrogenase complex repressor can be attenuated by deletion of the endogenous pyruvate dehydrogenase complex repressor gene.

**[0217]** In some cases, one or more genes of the pyruvate dehydrogenase complex are endogenous genes. Another way to increase the activity of the pyruvate dehydrogenase complex is by introducing into the microorganism one or more heterologous nucleic acids encoding one or more polypeptides from the group consisting of (a) pyruvate dehydrogenase (E1), (b) dihydrolipoyl transacetylase, and (c) dihydrolipoyl dehydrogenase.

**[0218]** By using any of these methods, the recombinant microorganism can produce increased amounts of acetyl Co-A in comparison to microorganisms wherein the activity of

pyruvate dehydrogenase is not modulated. Modulating the activity of pyruvate dehydrogenase can result in more carbon flux into the mevalonate dependent biosynthetic pathway in comparison to microorganisms that do not have modulated pyruvate dehydrogenase expression.

f. Combinations of mutations

**[0219]** It is understood that for any of the enzymes and/or enzyme pathways described herein, molecular manipulations that modulate any combination (two, three, four, five or six) of the enzymes and/or enzyme pathways described herein is expressly contemplated. For ease of the recitation of the combinations, citrate synthase (*gltA*) is designated as A, phosphotransacetylase (*ptaB*) is designated as B, acetate kinase (*ackA*) is designated as C, lactate dehydrogenase (*ldhA*) is designated as D, malic enzyme (*sfcA* or *maeB*) is designated as E, pyruvate decarboxylase (*aceE*, *aceF*, and/or *lpdA*) is designated as F, 6-phosphogluconolactonase *(ybhE)* is designated as G, and phosphoenolpyruvate carboxylase (*ppl*) is designated as H. As discussed above, aceE, aceF, and/or lpdA enzymes of the pyruvate decarboxylase complex can be used singly, or two of three enzymes, or three of three enzymes for increasing pyruvate decarboxylase activity.

**[0220]** Accordingly, for combinations of any two of the enzymes A-H, non-limiting combinations that can be used are: AB, AC, AD, AE, AF, AG, AH, BC, BD, BE, BF, BG, BH, CD, CE, CF, CG, CH, DE, DF, DG, DH, EF, EG, EH, and GH. For combinations of any three of the enzymes A-H, non-limiting combinations that can be used are: ABC, ABD, ABE, ABF, ABG, ABH, BCD, BCE, BCF, BCG, BCH, CDE, CDF, CDG, CDH, DEF, DEH, ACD, ACE, ACF, ACG, ACH, ADE, ADF, ADG, ADH, AEF, AEG, AEH, BDE, BDF, BDG, BDH, BEF, BEG, BEH, CEF, CEG, CEH, CFG, CFH, and CGH. For combinations of any four of the enzymes A-H, non-limiting combinations that can be used are: ABCD, ABCE, ABCF, ABCG, ABCH, ABDE, ABDF, ABDG, ABDH, ABEF, ABEG, ABEH, BCDE, BCDF, BCDG, BCDH, CDEF, CDEG, CDEH, ACDE, ACDF, ACDG, ACDH, ACEF, ACEG, ACEH, BCEF, BDEF, BGEF, BHEF, ADEF. For combinations of any five of the enzymes A-H, non-limiting combinations that can be used are: ABCDE, ABCDF, ABCDG, ABCDH, ABDEF, ABDEG, ABDEH, BCDEF, BCDEG, BCDEH, ACDEF, ACDEG, ACEDH, ABCEF, ABCEG, and ABCEH. For combinations of any six of the enzymes A-H, non-limiting combinations that can be used are: ABCDEF, ABCDEG, ABCDEH, BCDEFG, BCDEFH, and CDEFGH. For combinations of any seven of the enzymes A-H, non-limiting combinations that can be used are:

ABCDEFG, ABCDEFH, BCDEFGH. In another aspect, all eight enzyme combinations are used ABCDEFGH.

**[0221]** Accordingly, the recombinant microorganism as described herein can achieve increased mevalonate production that is increased compared to microorganisms that are not grown under conditions of tri-carboxylic acid (TCA) cycle activity, wherein metabolic carbon flux in the recombinant microorganism is directed towards mevalonate production by modulating the activity of one or more enzymes from the group consisting of (a) citrate synthase, (b) phosphotransacetylase; (c) acetate kinase; (d) lactate dehydrogenase; (e) NADP-dependent malic enzyme, and; (f) pyruvate dehydrogenase; (g) 6-phosphogluconolactonase; (h) phosphoenolpyruvate carboxylase; (i) the inhibitor of RssB activity during magnesium starvation protein; (j) the acrA component of the multidrug efflux pump acrAB-TolC; and (k) the fumarate and nitrate reduction sRNA (FNR).

## 7. Other regulators and factors for increased isoprene production

**[0222]** Other molecular manipulations can be used to increase the flow of carbon towards isoprene production. One method is to reduce, decrease or eliminate the effects of negative regulators for pathways that feed into the mevalonate pathway. For example, in some cases, the genes aceEF-lpdA are in an operon, with a fourth gene upstream pdhR. pdhR is a negative regulator of the transcription of its operon. In the absence of pyruvate, it binds its target promoter and represses transcription. It also regulates ndh and cyoABCD in the same way (Ogasawara, H. et al. 2007. J. Bact. 189:5534-5541). In one aspect, deletion of pdhR regulator can improve the supply of pyruvate, and hence the production of isoprene via the alternative lower MVA pathway (e.g., MVK, PMevDC, IPK, and/or IDI).

**[0223]** In other embodiments, any of the resultant strains described above can be further engineered to modulate the activity of the Entner-Doudoroff pathway. The gene coding for phosphogluconate dehydratase or aldolase can be attenuated or deleted. In other embodiments, any of the resultant strains described above may also be engineered to decrease or remove the activity of acetate kinase or citrate synthase. In other embodiments, any of the strains the resultant strain may also be engineered to decrease or remove the activity of phosphofructokinase. In other embodiments, any of the resultant strains described above may also be engineered to modulate the activity of glyceraldehyde-3-phosphate dehydrogenase. The

activity of glyceraldehyde-3-phosphate dehydrogenase can be modulated by decreasing its activity. In other embodiments, the enzymes from the non-oxidative branch of the pentose phosphate pathway, such as transketolase, transaldolase, ribulose-5-phosphate-epimerase and (or) ribose-5-phosphate epimerase can be overexpressed.

**[0224]** In other aspects, the host cells can be further engineered to increase intracellular acetyl-phospate concentrations by introducing heterologous nucleic acids encoding sedoheptulose-1,7-bisphosphatase/fructose-1,6 –bisphosphate aldolase and sedoheptulose-1,7-bisphosphatase/fructose-1,6–bisphosphate phosphatase. In certain embodiments, the host cells having these molecular manipulations can be combined with attenuated or deleted transaldolase (talB) and phosphofructokinase (pfkA and/or pfkB) genes, thereby allowing faster conversion of erythrose 4-phosphate, dihydroxyacetone phosphate, and glyceraldehyde 3-phosphate into sedoheptulose 7-phosphate and fructose 1-phosphate.

**[0225]** In other aspects, the introduction of 6-phosphogluconolactonase (PGL) into microorganisms (such as various *E. coli* strains) which lack PGL can be used to improve production of isoprene. PGL may be introduced using chromosomal integration or extra-chromosomal vehicles, such as plasmids. In yet other aspects, PGL may be deleted from the genome of cells (for example, microorganisms, such as various *E. coli* strains) which express a PGL to improve production of mevalonate and/or isoprene. In another aspect, a heterologous nucleic acid encoding a PGL polypeptide can be expressed in a cell which does not endogenously express PGL. In some aspects, deletion of PGL results in any of about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100%, inclusive, such as any values in between these percentages, higher percent yield of isoprene in comparison to microorganisms that express PGL. In other aspects, deletion of PGL results in any of about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100%, inclusive, including any values in between these percentages, higher instantaneous percent yield of isoprene in comparison to microorganisms that express PGL. In other aspects, deletion of PGL results in any of about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100%, inclusive, including any values in between these percentages, higher cell productivity index for isoprene in comparison to microorganisms that express PGL. In other aspects, deletion of PGL results in any of about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100%, inclusive, including any values in between these percentages, higher volumetric productivity of isoprene in comparison to microorganisms that

express PGL. In other aspects, deletion of PGL results in any of about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100%, inclusive, including any values in between these percentages, higher peak specific productivity of isoprene in comparison to microorganisms that express PGL. In some aspects the deletion of PGL results in peak specific productivity being maintained for a longer period of time in comparison to microorganisms that express PGL.

## 8. Vectors

**[0226]** Suitable vectors can be used for any of the compositions and methods described herein. For example, suitable vectors can be used to optimize the expression of one or more copies of a gene encoding an isoprene synthase, an acetoacetyl co-A synthase, an MVA pathway enzyme, a DXP pathway enzyme, a phosphoketolase, and/or a polyprenyl pyrophosphate synthase, in a cell. In some aspects, the vector contains a selective marker. Examples of selectable markers include, but are not limited to, antibiotic resistance nucleic acids (*e.g.*, kanamycin, ampicillin, carbenicillin, gentamicin, hygromycin, phleomycin, bleomycin, neomycin, or chloramphenicol) and/or nucleic acids that confer a metabolic advantage, such as a nutritional advantage on the host cell. In some aspects, one or more copies of an an isoprene synthase, an acetoacetyl co-A synthase, an MVA pathway enzyme, a DXP pathway enzyme, a phosphoketolase, and/or a polyprenyl pyrophosphate synthases nucleic acid(s) integrate into the genome of host cells without a selective marker.

**[0227]** Any one of the vectors characterized or used in the Examples of the present disclosure can be used.

## 9. Exemplary host cells

**[0228]** One of skill in the art will recognize that expression vectors are designed to contain certain components which optimize gene expression for certain host strains. Such optimization components include, but are not limited to origin of replication, promoters, and enhancers. The vectors and components referenced herein are described for exemplary purposes and are not meant to narrow the scope of the invention.

**[0229]** Any microorganism or progeny thereof that can be used to heterologously express genes can be used for modulation of any of the genes described herein for increased production of isoprene (*e.g.*, citrate synthase, phosphotransacetylase, acetate kinase, lactate dehydrogenase,

malic enzyme, pyruvate dehydrogenase, 6-phosphogluconolactonase, phosphoenolpyruvate carboxylase and/or FNR). Also, any microorganism or progeny thereof that can be used to heterologously express genes can be used to express one or more heterologous nucleic acids encoding an isoprene synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptides in a cell. In some aspects, any microorganism or progeny thereof can be used to heterologously express one or more genes encoding isoprene synthase (such as, but not limited to, isoprene synthase genes from *A. hypogaea*, *M. pruriens*, *C. cajans*, *G. max*, *G. soja*, *L. japonicus*, *M. truncatula*, *C. limon*, *C. tenuipile*, *V. vinifera*, *Q. petraea*, *Q. ilex*, *E. globulus*, and/or *M. alternifolia*). Bacteria cells, including gram positive or gram negative bacteria can be used to express any of the nucleic acids or polypeptides described above. In some embodiments, the host cell is a gram-positive bacterium. Non-limiting examples include strains of *Streptomyces* (e.g., *S. lividans, S. coelicolor, S. rubiginosus*, or *S. griseus*), *Streptococcus*, *Bacillus* (e.g., *B. lichenformis* or *B. subtilis*), *Listeria* (e.g., *L. monocytogenes*), *Corynebacteria*, or *Lactobacillus* (e.g., *L. spp*). In some embodiments, the source organism is a gram-negative bacterium. Non-limiting examples include strains of *Escherichia* (*e.g., E. coli*), *Pseudomonas* (*e.g., P. alcaligenes*), *Pantoea* (e.g., *P. citrea*), *Enterobacter*, or *Helicobacter* (*H. pylori*). In particular, one or more copies of a nucleic acid encoding an isoprene synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptides can be expressed in any one of *P. citrea, B. subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. clausii, B. halodurans, B. megaterium, B. coagulans, B. circulans, B. lautus, B. thuringiensis, S. albus, S. lividans, S. coelicolor, S. griseus, Pseudomonas sp.,* and *P. alcaligenes* cells. In some aspects, the host cell can be a *Lactobacilis spp.,* such as *Lactobacillus lactis* or a *Lactobacillus plantarum.*

**[0230]** Exemplary host cells include, for example, yeasts, such as species of *Saccharomyces* (*e.g., S. cerevisiae*), bacteria, such as species of *Escherichia* (*e.g., E. coli*), archaea, such as species of *Methanosarcina* (*e.g., Methanosarcina mazei*), plants, such as kudzu or poplar (*e.g., Populus alba* or *Populus alba x tremula* CAC35696) or aspen (*e.g., Populus tremuloides*).

**[0231]** There are numerous types of anaerobic cells that can be used as host cells in the compositions and methods of the present invention. In one aspect of the invention, the cells described in any of the compositions or methods described herein are obligate anaerobic cells and progeny thereof. Obligate anaerobes typically do not grow well, if at all, in conditions where oxygen is present. It is to be understood that a small amount of oxygen may be present, that is, there is some tolerance level that obligate anaerobes have for a low level of oxygen. In one aspect, obligate anaerobes engineered to produce isoprene can serve as host cells for any of the methods and/or compositions described herein and are grown under substantially oxygen-free conditions, wherein the amount of oxygen present is not harmful to the growth, maintenance, and/or fermentation of the anaerobes.

**[0232]** In another aspect of the invention, the host cells described and/or used in any of the compositions or methods described herein are facultative anaerobic cells and progeny thereof. Facultative anaerobes can generate cellular ATP by aerobic respiration (*e.g.*, utilization of the TCA cycle) if oxygen is present. However, facultative anaerobes can also grow in the absence of oxygen. This is in contrast to obligate anaerobes which die or grow poorly in the presence of greater amounts of oxygen. In one aspect, therefore, facultative anaerobes can serve as host cells for any of the compositions and/or methods provided herein and can be engineered to produce isoprene. Facultative anaerobic host cells can be grown under substantially oxygen-free conditions, wherein the amount of oxygen present is not harmful to the growth, maintenance, and/or fermentation of the anaerobes, or can be alternatively grown in the presence of greater amounts of oxygen.

**[0233]** The host cell can additionally be a filamentous fungal cell and progeny thereof. (*See, e.g.,* Berka & Barnett, *Biotechnology Advances,* (1989), 7(2):127-154). In some aspects, the filamentous fungal cell can be any of *Trichoderma longibrachiatum, T. viride, T. koningii, T. harzianum, Penicillium sp., Humicola insolens, H. lanuginose, H. grisea, Chrysosporium sp., C. lucknowense, Gliocladium sp., Aspergillus sp.*, such as *A. oryzae, A. niger, A sojae, A. japonicus, A. nidulans*, or *A. awamori, Fusarium sp.,* such as *F. roseum, F. graminum F. cerealis, F. oxysporuim*, or *F. venenatum, Neurospora sp.,*such as *N. crassa, Hypocrea sp., Mucor sp.,*such as *M. miehei, Rhizopus sp.* or *Emericella sp.* In some aspects, the fungus is *A. nidulans, A. awamori, A. oryzae, A. aculeatus, A. niger, A. japonicus, T. reesei, T. viride, F. oxysporum,* or *F.*

*solani*. In certain embodiments, plasmids or plasmid components for use herein include those described in U.S. patent pub. No. US 2011/0045563.

**[0234]** The host cell can also be a yeast, such as *Saccharomyces sp., Schizosaccharomyces sp., Pichia sp.,* or *Candida sp*. In some aspects, the *Saccharomyces sp.* is *Saccharomyces cerevisiae* (See, *e.g.*, Romanos et al., *Yeast*, (1992), 8(6):423-488). In certain embodiments, plasmids or plasmid components for use herein include those described in U.S. pat. No, 7,659,097 and U.S. patent pub. No. US 2011/0045563.

**[0235]** The host cell can also be a species of plant, such as a plant from the family Fabaceae, such as the Faboideae subfamily. In some aspects, the host cell is kudzu, poplar (such as *Populus alba x tremula* CAC35696), aspen (such as *Populus tremuloides*), or *Quercus robur*.

**[0236]** The host cell can additionally be a species of algae, such as a green algae, red algae, glaucophytes, chlorarachniophytes, euglenids, chromista, or dinoflagellates. (See, *e.g.*, Saunders & Warmbrodt, "*Gene Expression in Algae and Fungi, Including Yeast*," (1993), National Agricultural Library, Beltsville, MD). In certain embodiments, plasmids or plasmid components for use herein include those described in U.S. Patent Pub. No. US 2011/0045563. In some aspects, the host cell is a cyanobacterium, such as cyanobacterium classified into any of the following groups based on morphology: *Chlorococcales, Pleurocapsales, Oscillatoriales, Nostocales,* or *Stigonematales* (See, *e.g.*, Lindberg et al., Metab. Eng., (2010) 12(1):70-79). In certain embodiments, plasmids or plasmid components for use herein include those described in U.S. Patent Pub. No.: US 2010/0297749; US 2009/0282545 and Intl. Pat. Appl. No. WO 2011/034863.

**[0237]** *E. coli* host cells can be used to express one or more isoprene synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptides in the compositions and methods described herein. In one aspect, the host cell is a recombinant cell of an *Escherichia coli* (*E. coli*) strain, or progeny thereof, capable of producing isoprene, that expresses one or more nucleic acids encoding isoprene synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptide. The *E. coli* host cells can produce isoprene in amounts, peak titers, and cell productivities greater than that of the same cells lacking one or more heterologously expressed nucleic acids encoding isoprene

synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptide. In addition, the one or more heterologously expressed nucleic acids encoding isoprene synthase, acetoacetyl co-A synthase, MVA pathway enzyme, DXP pathway enzyme, phosphoketolase, and/or a polyprenyl pyrophosphate synthase polypeptide in *E. coli* can be chromosomal copies (*e.g.*, integrated into the *E. coli* chromosome). In other aspects, the *E. coli* cells are in culture.

**[0238]** In other aspects, the host cell can be a species of yeast other than *S. cerevisiae* such as, but not limited to, a *Pichia spp.,* a *Candida spp.,* a *Hansenula spp.,* a *Kluyveromyces spp.,* a *Kluyveromyces spp.,* or a *Schizosaccharomyces spp.* In still other aspects, the host cell can be a species of bacterium including, but not limited to, an *Arthrobacter spp.,* a *Zymomonas spp.,* a *Brevibacterium spp.,* a *Clostridium spp.,* an *Aerococcus spp.,* a *Bacillus spp.,* an *Actinobacillus spp.* (such as, but not limited to, *A. succinogens*), a *Carbobacterium spp.,* a *Corynebacterium spp.,* an *Enterococcus spp.,* an *Erysipelothrix spp.,* a *Gemella spp.,* a *Geobacillus spp.,* a *Globicatella spp.,* a *Lactobacillus spp.*(such as, but not limited to, *L. lactis* and *L. rhammosus*), a *Lactococcus spp.,* a *Leuconostoc spp.,* a *Pediococcus spp.,* a *Streptococcus spp.,* a *Tetragenococcus spp.,* an *Actinobacillus spp.,* or a *Vagococcus spp.,* In other aspects, the fermenting organism can be a fungus such as, but not limited to, a *Rhizopus spp.*

**[0239]** In other aspects, the host cell can be a lactic acid bacteria, such as those of the genera Aerococcus, Bacillus, Carbobacterium, Enterococcus, Erysipelothrix, Gemella, Globicatella, Lactobacillus, Lactococcus, Leuconostoc, Pediococcus, Streptococcus, Tetragenococcus and Vagococcus. For example, other bacteria of the genus Lactobacillus which may be substituted include, but are not limited to, L. heiveticus, L. delbrueckii, L. casei, L, acidophilus, L. amylovorus, L. leichmanii or L. bulgaricus. L. amylovorus, and L. pentosus.

**[0240]** Other exemplary host cells that can be used are described in US Pub. 2009/0203102, WO 2009/076676, WO 2010/003007, WO 2009/132220, WO 2010/031062, WO 2010/031068, WO 2010/031076, WO 2010/031077, and WO 2010/031079.

## B. Methods for the production of isoprene

**[0241]** Provided herein are methods of producing isoprene by culturing any of the recombinant cells described herein under conditions such as those disclosed herein using any

isoprene as described herein. In one aspect, isoprene can be produced by culturing recombinant cells expressing one or more nucleic acids encoding: (a) an isoprene synthase polypeptide, wherein the isoprene synthase polypeptide is encoded by a heterologous nucleic acid; and (b) one or more mevalonate (MVA) pathway polypeptides in culture media. In one aspect, one or more heterologous nucleic acids encoding a thiolase, HMG-CoA reductase, a lower MVA pathway polypeptide, and an isoprene synthase polypeptide can be used. In another aspect, isoprene can be produced by culturing recombinant cells comprising one or more heterologous nucleic acids encoding a thiolase, HMG-CoA reductase and HMG-CoA synthase, a lower MVA pathway polypeptide, and an isoprene synthase polypeptide. In yet another aspect, one or more heterologous nucleic acids encoding one or more upper MVA pathway polypeptides, one or more lower MVA pathway polypeptides, and/or one or more DXP pathway polypeptides can be used. In some aspects, the recombinant cells described herein exhibit any of about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 95%, or 100%, inclusive, including any value in between these percentages, increased isoprene production in comparison to cells which do not comprise one or more heterologous nucleic acids encoding an isoprene synthase polypeptide and one or more MVA pathway polypeptides. The isoprene can be produced from any of the cells described herein and according to any of the methods described herein. Any of the cells can be used for the purpose of producing isoprene using any isoprene synthase as described herein.

**[0242]** The cells can further express one or more nucleic acid molecules encoding the lower MVA pathway polypeptide(s) described above (*e.g.*, MVK, PMK, MVD, and/or IDI), any of the upper MVA pathways polypeptide(s) described above (*e.g.,* a thiolase, an acetoacetyl-CoA synthase, an HMG-CoA reductase, and/or an HMG-CoA synthase) and/or any of the isoprene synthase polypeptide(s) described above (*e.g. A. hypogaea* isoprene synthase). In some aspects, the recombinant (*e.g.,* bacterial) cells can be any of the cells described herein. Any of the isoprene synthases or variants thereof described herein, any of the bacterial strains described herein, any of the promoters described herein, and/or any of the vectors described herein can also be used to produce isoprene. In some aspects, the method of producing isoprene further comprises a step of recovering the isoprene.

**[0243]** In some aspects, the amount of isoprene produced is measured at a productivity time point. In some aspects, the productivity for the cells is about any of the amounts of isoprene

disclosed herein. In some aspects, the cumulative, total amount of isoprene produced is measured. In some aspects, the cumulative total productivity for the cells is about any of the amounts of isoprene disclosed herein.

**[0244]** In some aspects, any of the cells described herein (for examples the cells in culture) produce isoprene at greater than about any of or about any of 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, or more nmole of isoprene/gram of cells for the wet weight of the cells/hour (nmole/$g_{wcm}$/hr). In some aspects, the amount of isoprene is between about 2 to about 5,000 nmole/$g_{wcm}$/hr, such as between about 2 to about 100 nmole/$g_{wcm}$/hr, about 100 to about 500 nmole/$g_{wcm}$/hr, about 150 to about 500 nmole/$g_{wcm}$ /hr, about 500 to about 1,000 nmole/$g_{wcm}$/hr, about 1,000 to about 2,000 nmole/$g_{wcm}$/hr, or about 2,000 to about 5,000 nmole/$g_{wcm}$/hr. In some aspects, the amount of isoprene is between about 20 to about 5,000 nmole/$g_{wcm}$/hr, about 100 to about 5,000 nmole/$g_{wcm}$/hr, about 200 to about 2,000 nmole/$g_{wcm}$/hr, about 200 to about 1,000 nmole/$g_{wcm}$/hr, about 300 to about 1,000 nmole/$g_{wcm}$/hr, or about 400 to about 1,000 nmole/$g_{wcm}$/hr.

**[0245]** In some aspects, the cells in culture produce isoprene at greater than or about 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, 10,000, 100,000, or more ng of isoprene/gram of cells for the wet weight of the cells/hr (ng/$g_{wcm}$/h). In some aspects, the amount of isoprene is between about 2 to about 5,000 ng/$g_{wcm}$/h, such as between about 2 to about 100 ng/$g_{wcm}$/h, about 100 to about 500 ng/$g_{wcm}$/h, about 500 to about 1,000 ng/$g_{wcm}$/h, about 1,000 to about 2,000 ng/$g_{wcm}$/h, or about 2,000 to about 5,000 ng/$g_{wcm}$/h. In some aspects, the amount of isoprene is between about 20 to about 5,000 ng/$g_{wcm}$/h, about 100 to about 5,000 ng/$g_{wcm}$/h, about 200 to about 2,000 ng/$g_{wcm}$/h, about 200 to about 1,000 ng/$g_{wcm}$/h, about 300 to about 1,000 ng/$g_{wcm}$/h, or about 400 to about 1,000 ng/$g_{wcm}$/h.

**[0246]** In some aspects, the cells in culture produce a cumulative titer (total amount) of isoprene at greater than about any of or about any of 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, 10,000, 50,000, 100,000, or more mg of isoprene/L of broth (mg/$L_{broth}$, wherein the volume of broth includes the volume of the cells and the cell medium). In some aspects, the amount of isoprene is between about 2 to about 5,000 mg/$L_{broth}$, such as between about 2 to about 100 mg/$L_{broth}$, about 100 to about 500 mg/$L_{broth}$, about 500 to about 1,000 mg/$L_{broth}$, about 1,000 to about 2,000

mg/$L_{broth}$, or about 2,000 to about 5,000 mg/$L_{broth}$. In some aspects, the amount of isoprene is between about 20 to about 5,000 mg/$L_{broth}$, about 100 to about 5,000 mg/$L_{broth}$, about 200 to about 2,000 mg/$L_{broth}$, about 200 to about 1,000 mg/$L_{broth}$, about 300 to about 1,000 mg/$L_{broth}$, or about 400 to about 1,000 mg/$L_{broth}$.

**[0247]** In some aspects, the isoprene produced by the cells in culture (such as any of the recombinant cells described herein) comprises at least about 1, 2, 5, 10, 15, 20, or 25% by volume of the fermentation offgas. In some aspects, the isoprene comprises between about 1 to about 25% by volume of the offgas, such as between about 5 to about 15 %, about 15 to about 25%, about 10 to about 20%, or about 1 to about 10 %.

**[0248]** In some embodiments, the cells are cultured in a culture medium under conditions permitting the production of isoprene by the cells. In some embodiments, the cells in culture produce isoprene at greater than or about 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, or more nmole of isoprene/gram of cells for the wet weight of the cells/hour (nmole/gwcm/hr). In some embodiments, the amount of isoprene is between about 2 to about 5,000 nmole/gwcm/hr, such as between about 2 to about 100 nmole/gwcm/hr, about 100 to about 500 nmole/gwcm/hr, about 150 to about 500 nmole/gwcm /hr, about 500 to about 1,000 nmole/gwcm/hr, about 1,000 to about 2,000 nmole/gwcm/hr, or about 2,000 to about 5,000 nmole/gwcm/hr. The amount of isoprene in units of nmole/gwcm/hr can be measured as disclosed in U.S. Patent No. 5,849,970. For example, two mL of headspace (*e.g.*, headspace from a culture such as 2 mL of culture cultured in sealed vials at 32°C with shaking at 200 rpm for approximately 3 hours) are analyzed for isoprene using a standard gas chromatography system, such as a system operated isothermally (85°C) with an n-octane/porasil C column (Alltech Associates, Inc., Deerfield, IL) and coupled to a RGD2 mercuric oxide reduction gas detector (Trace Analytical, Menlo Park, CA) (*see*, *e.g.*, Greenberg et al, Atmos. Environ. 27A: 2689-2692, 1993; Silver et al., Plant Physiol. 97:1588-1591, 1991). The gas chromatography area units are converted to nmol isoprene via a standard isoprene concentration calibration curve. In some embodiments, the value for the grams of cells for the wet weight of the cells is calculated by obtaining the $A_{600}$ value for a sample of the cell culture, and then converting the $A_{600}$ value to grams of cells based on a calibration curve of wet weights for cell cultures with a known $A_{600}$ value. In some embodiments, the grams of the cells is estimated by assuming that one liter of broth (including

cell medium and cells) with an $A_{600}$ value of 1 has a wet cell weight of 1 gram. The value is also divided by the number of hours the culture has been incubating for, such as three hours.

**[0249]** In some embodiments, the cells in culture produce isoprene at greater than or about 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, 10,000, 100,000, or more ng of isoprene/gram of cells for the wet weight of the cells/hr (ng/gwcm/h). In some embodiments, the amount of isoprene is between about 2 to about 5,000 ng/gwcm/h, such as between about 2 to about 100 ng/gwcm/h, about 100 to about 500 ng/gwcm/h, about 500 to about 1,000 ng/gwcm/h, about 1,000 to about 2,000 ng/gwcm/h, or about 2,000 to about 5,000 ng/gwcm/h. The amount of isoprene in ng/gwcm/h can be calculated by multiplying the value for isoprene production in the units of nmole/gwcm/hr discussed above by 68.1 (as described in Equation 5 below).

**[0250]** In some embodiments, the cells in culture produce a cumulative titer (total amount) of isoprene at greater than or about 1, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,250, 1,500, 1,750, 2,000, 2,500, 3,000, 4,000, 5,000, 10,000, 50,000, 100,000, or more mg of isoprene/L of broth (mg/L broth, wherein the volume of broth includes the volume of the cells and the cell medium). In some embodiments, the amount of isoprene is between about 2 to about 5,000 mg/L broth, such as between about 2 to about 100 mg/L broth, about 100 to about 500 mg/L broth, about 500 to about 1,000 mg/L broth, about 1,000 to about 2,000 mg/L broth, or about 2,000 to about 5,000 mg/L broth. The specific productivity of isoprene in mg of isoprene/L of headspace from shake flask or similar cultures can be measured by taking a 1 ml sample from the cell culture at an $OD_{600}$ value of approximately 1.0, putting it in a 20 mL vial, incubating for 30 minutes, and then measuring the amount of isoprene in the headspace. If the $OD_{600}$ value is not 1.0, then the measurement can be normalized to an $OD_{600}$ value of 1.0 by dividing by the $OD_{600}$ value. The value of mg isoprene/L headspace can be converted to mg/Lbroth/hr/$OD_{600}$ of culture broth by multiplying by a factor of 38. The value in units of mg/Lbroth/hr/$OD_{600}$ can be multiplied by the number of hours and the $OD_{600}$ value to obtain the cumulative titer in units of mg of isoprene/L of broth.

**[0251]** The instantaneous isoprene production rate in mg/L broth/hr in a fermentor can be measured by taking a sample of the fermentor off-gas, analyzing it for the amount of isoprene (in units such as mg of isoprene per L of gas), and multiplying this value by the rate at which off-gas is passed though each liter of broth (*e.g.*, at 1 vvm (volume of air/volume of broth/minute)

this is 60 Lgas per hour). Thus, an off-gas level of 1 mg/Lgas corresponds to an instantaneous production rate of 60 mg/Lbroth/hr at air flow of 1 vvm. If desired, the value in the units mg/Lbroth/hr can be divided by the $OD_{600}$ value to obtain the specific rate in units of mg/Lbroth/hr/OD. The average value of mg isoprene/Lgas can be converted to the total product productivity (grams of isoprene per liter of fermentation broth, mg/Lbroth) by multiplying this average off-gas isoprene concentration by the total amount of off-gas sparged per liter of fermentation broth during the fermentation. Thus, an average off-gas isoprene concentration of 0.5 mg/Lbroth/hr over 10 hours at 1 vvm corresponds to a total product concentration of 300 mg isoprene/Lbroth.

**[0252]** In some embodiments, the cells in culture convert greater than or about 0.0015, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.12, 0.14, 0.16, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.4, or 1.6% of the carbon in the cell culture medium into isoprene. In some embodiments, the percent conversion of carbon into isoprene is between about 0.002 to about 1.6%, such as about 0.002 to about 0.005%, about 0.005 to about 0.01%, about 0.01 to about 0.05%, about 0.05 to about 0.15%, 0.15 to about 0.2%, about 0.2 to about 0.3%, about 0.3 to about 0.5%, about 0.5 to about 0.8%, about 0.8 to about 1.0%, or about 1.0 to about 1.6%. The percent conversion of carbon into isoprene (also referred to as "% carbon yield") can be measured by dividing the moles carbon in the isoprene produced by the moles carbon in the carbon source (such as the moles of carbon in batched and fed glucose and yeast extract). This number is multiplied by 100% to give a percentage value (as indicated in Equation 1).

**Equation 1**

% Carbon Yield = (moles carbon in isoprene produced)/(moles carbon in carbon source) * 100

**[0253]** For this calculation, yeast extract can be assumed to contain 50% w/w carbon.

**Equation 2**

% Carbon Yield = (39.1 g isoprene * 1/68.1mol/g * 5 C/mol)/[(181221 g glucose * 1/180 mol/g * 6 C/mol) + (17780 g yeast extract * 0.5 * 1/12 mol/g)] * 100 = 0.042%

**[0254]** One skilled in the art can readily convert the rates of isoprene production or amount of isoprene produced into any other units. Exemplary equations are listed below for interconverting between units.

**[0255]** Units for Rate of Isoprene production (total and specific)

**Equation 3**

1 g isoprene/$L_{broth}$/hr = 14.7 mmol isoprene/$L_{broth}$/hr (total volumetric rate)

**Equation 4**

1 nmol isoprene /$g_{wcm}$/hr = 1 nmol isoprene /$L_{broth}$/hr/$OD_{600}$ (This conversion assumes that one liter of broth with an $OD_{600}$ value of 1 has a wet cell weight of 1 gram.)

**Equation 5**

1 nmol isoprene/$g_{wcm}$/hr = 68.1 ng isoprene/$g_{wcm}$/hr (given the molecular weight of isoprene)

**Equation 6**

1 nmol isoprene/$L_{gas}$ $O_2$/hr = 90 nmol isoprene/$L_{broth}$/hr (at an $O_2$ flow rate of 90 L/hr per L of culture broth)

**Equation 7**

1 μg isoprene/$L_{gas}$ isoprene in off-gas = 60 μg isoprene/$L_{broth}$/hr at a flow rate of 60 $L_{gas}$ per $L_{broth}$ (1 vvm)

**[0256]** Units for Titer (total and specific)

**Equation 8**

1 nmol isoprene/mg cell protein = 150 nmol isoprene/$L_{broth}$/$OD_{600}$ (This conversion assumes that one liter of broth with an $OD_{600}$ value of

1 has a total cell protein of approximately 150 mg) (specific productivity)

**Equation 9**

1 g isoprene/$L_{broth}$ = 14.7 mmol isoprene/$L_{broth}$ (total titer)

**[0257]** If desired, Equation 10 can be used to convert any of the units that include the wet weight of the cells into the corresponding units that include the dry weight of the cells.

**Equation 10**

Dry weight of cells = (wet weight of cells)/3.3

**[0258]** In some embodiments encompassed by the invention, a cell comprising a heterologous nucleic acid encoding an isoprene synthase polypeptide produces an amount of isoprene that is at least or about 2-fold, 3-fold, 5-fold, 10-fold, 25-fold, 50-fold, 100-fold, 150-fold, 200-fold, 400-fold, or greater than the amount of isoprene produced from a corresponding cell grown under essentially the same conditions without the heterologous nucleic acid encoding the isoprene synthase polypeptide.

**[0259]** In some embodiments encompassed by the invention, a cell comprising a heterologous nucleic acid encoding an isoprene synthase polypeptide and one or more heterologous nucleic acids encoding a DXS, IDI, and/or MVA pathway polypeptide produces an amount of isoprene that is at least or about 2-fold, 3-fold, 5-fold, 10-fold, 25-fold, 50-fold, 100-fold, 150-fold, 200-fold, 400-fold, or greater than the amount of isoprene produced from a corresponding cell grown under essentially the same conditions without the heterologous nucleic acids.

## 1. Transformation methods

**[0260]** Nucleic acids encoding one or more copies of a nucleic acid encoding an mvaE and an mvaS polypeptide, an isoprene synthase polypeptide, MVA pathway polypeptides, DXP pathway polypeptides, phosphoketolase polypeptide, and/or polyprenyl pyrophosphate synthase polypeptides can be inserted into a microorganism using suitable techniques. Additionally, these nucleic acids or vectors containing them can be inserted into a host cell (*e.g.*, a plant cell, a

fungal cell, a yeast cell, or a bacterial cell described herein) using standard techniques for introduction of a DNA construct or vector into a host cell, such as transformation, electroporation, nuclear microinjection, transduction, transfection (*e.g.*, lipofection mediated or DEAE-Dextrin mediated transfection or transfection using a recombinant phage virus), incubation with calcium phosphate DNA precipitate, high velocity bombardment with DNA-coated microprojectiles, and protoplast fusion. General transformation techniques are known in the art (*See, e.g., Current Protocols in Molecular Biology* (F. M. Ausubel *et al.* (eds.) Chapter 9, 1987; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor, 1989; and Campbell et al., Curr. Genet. 16:53-56, 1989). The introduced nucleic acids can be integrated into chromosomal DNA or maintained as extrachromosomal replicating sequences. Transformants can be selected by any method known in the art. Suitable methods for selecting transformants are described in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (US Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, and US Publ. No. 2010/0003716.

## 2. Exemplary Cell Culture Media

**[0261]** As used herein, the terms "minimal medium" or "minimal media" refer to growth medium containing the minimum nutrients possible for cell growth, generally, but not always, without the presence of amino acids (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more amino acids). Minimal medium can contain: (1) a carbon source for microbial growth; (2) various salts, which can vary among microbial species and growing conditions; and (3) water. The carbon source can vary significantly, from simple sugars like glucose to more complex hydrolysates of other biomass, such as yeast extract, as discussed in more detail below. The salts generally provide essential elements such as magnesium, nitrogen, phosphorus, and sulfur to allow the cells to synthesize proteins and nucleic acids. Minimal medium can also be supplemented with selective agents, such as antibiotics, to select for the maintenance of certain plasmids and the like. For example, if a microorganism is resistant to a certain antibiotic, such as ampicillin or tetracycline, then that antibiotic can be added to the medium in order to prevent cells lacking the resistance from growing. Medium can be supplemented with other compounds as necessary to select for desired physiological or biochemical characteristics, such as particular amino acids and the like.

**[0262]** Any minimal medium formulation can be used to cultivate the host cells. Exemplary minimal medium formulations include, for example, M9 minimal medium and TM3 minimal

medium. Each liter of M9 minimal medium contains (1) 200 ml sterile M9 salts (64 g $Na_2HPO_4$-$7H_2O$, 15 g $KH_2PO_4$, 2.5 g NaCl, and 5.0 g $NH_4Cl$ per liter); (2) 2 ml of 1 M $MgSO_4$ (sterile); (3) 20 ml of 20% (w/v) glucose (or other carbon source); and (4) 100 µl of 1 M $CaCl_2$ (sterile). Each liter of TM3 minimal medium contains (1) 13.6 g $K_2HPO_4$; (2) 13.6 g $KH_2PO_4$; (3) 2 g $MgSO_4*7H_2O$; (4) 2 g Citric Acid Monohydrate; (5) 0.3 g Ferric Ammonium Citrate; (6) 3.2 g $(NH_4)_2SO_4$; (7) 0.2 g yeast extract; and (8) 1 ml of 1000X Trace Elements solution; pH is adjusted to ~6.8 and the solution is filter sterilized. Each liter of 1000X Trace Elements contains: (1) 40 g Citric Acid Monohydrate; (2) 30 g $MnSO_4*H_2O$; (3) 10 g NaCl; (4) 1 g $FeSO_4*7H_2O$; (4)1 g $CoCl_2*6H_2O$; (5) 1 g $ZnSO_4*7H_2O$; (6) 100 mg $CuSO_4*5H_2O$; (7) 100 mg $H_3BO_3$; and (8) 100 mg $NaMoO_4*2H_2O$; pH is adjusted to ~3.0.

**[0263]** An additional exemplary minimal media includes (1) potassium phosphate $K_2HPO_4$, (2) Magnesium Sulfate $MgSO_4$ * $7H_2O$, (3) citric acid monohydrate $C_6H_8O_7*H_2O$, (4) ferric ammonium citrate $NH_4FeC_6H_5O_7$, (5) yeast extract (from biospringer), (6) 1000X Modified Trace Metal Solution, (7) sulfuric acid 50% w/v, (8) foamblast 882 (Emerald Performance Materials), and (9) Macro Salts Solution 3.36ml. All of the components are added together and dissolved in deionized $H_2O$ and then heat sterilized. Following cooling to room temperature, the pH is adjusted to 7.0 with ammonium hydroxide (28%) and q.s. to volume. Vitamin Solution and spectinomycin are added after sterilization and pH adjustment.

**[0264]** Any carbon source can be used to cultivate the host cells. The term "carbon source" refers to one or more carbon-containing compounds capable of being metabolized by a host cell or organism. For example, the cell medium used to cultivate the host cells can include any carbon source suitable for maintaining the viability or growing the host cells. In some aspects, the carbon source is a carbohydrate (such as monosaccharide, disaccharide, oligosaccharide, or polysaccharides), or invert sugar (*e.g.*, enzymatically treated sucrose syrup).

**[0265]** In some aspects, the carbon source includes yeast extract or one or more components of yeast extract. In some aspects, the concentration of yeast extract is 0.1% (w/v), 0.09% (w/v), 0.08% (w/v), 0.07% (w/v), 0.06% (w/v), 0.05% (w/v), 0.04% (w/v), 0.03% (w/v), 0.02% (w/v), or 0.01% (w/v) yeast extract. In some aspects, the carbon source includes both yeast extract (or one or more components thereof) and another carbon source, such as glucose.

**[0266]** Exemplary monosaccharides include glucose and fructose; exemplary oligosaccharides include lactose and sucrose, and exemplary polysaccharides include starch and cellulose. Exemplary carbohydrates include C6 sugars (*e.g.*, fructose, mannose, galactose, or glucose) and C5 sugars (*e.g.*, xylose or arabinose).

**[0267]** In some aspects, the cells described herein are capable of using syngas as a source of energy and/or carbon. In some embodiments, the syngas includes at least carbon monoxide and hydrogen. In some embodiments, the syngas further additionally includes one or more of carbon dioxide, water, or nitrogen. In some embodiments, the molar ratio of hydrogen to carbon monoxide in the syngas is 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 3.0, 4.0, 5.0, or 10.0. In some embodiments, the syngas comprises 10, 20, 30, 40, 50, 60, 70, 80, or 90% by volume carbon monoxide. In some embodiments, the syngas comprises 10, 20, 30, 40, 50, 60, 70, 80, or 90% by volume hydrogen. In some embodiments, the syngas comprises 10, 20, 30, 40, 50, 60, 70, 80, or 90% by volume carbon dioxide. In some embodiments, the syngas comprises 10, 20, 30, 40, 50, 60, 70, 80, or 90% by volume water. In some embodiments, the syngas comprises 10, 20, 30, 40, 50, 60, 70, 80, or 90% by volume nitrogen.

**[0268]** Synthesis gas may be derived from natural or synthetic sources. The source from which the syngas is derived is referred to as a "feedstock." In some embodiments, the syngas is derived from biomass (e.g., wood, switch grass, agriculture waste, municipal waste) or carbohydrates (e.g., sugars). In other embodiments, the syngas is derived from coal, petroleum, kerogen, tar sands, oil shale, or natural gas. In other embodiments, the syngas is derived from rubber, such as from rubber tires.

**[0269]** Syngas can be derived from a feedstock by a variety of processes, including methane reforming, coal liquefaction, co-firing, fermentative reactions, enzymatic reactions, and biomass gasification. Biomass gasification is accomplished by subjecting biomass to partial oxidation in a reactor at temperatures above about 700 ºC in the presence of less than a stoichiometric amount of oxygen. The oxygen is introduced into the bioreactor in the form of air, pure oxygen, or steam. Gasification can occur in three main steps: 1) initial heating to dry out any moisture embedded in the biomass; 2) pyrolysis, in which the biomass is heated to 300-500 ºC in the absence of oxidizing agents to yield gas, tars, oils and solid char residue; and 3) gasification of solid char, tars and gas to yield the primary components of syngas. Co-firing is accomplished by

gasification of a coal/biomass mixture. The composition of the syngas, such as the identity and molar ratios of the components of the syngas, can vary depending on the feedstock from which it is derived and the method by which the feedstock is converted to syngas.

**[0270]** Synthesis gas can contain impurities, the nature and amount of which vary according to both the feedstock and the process used in production. Fermentations may be tolerant to some impurities, but there remains the need to remove from the syngas materials such as tars and particulates that might foul the fermentor and associated equipment. It is also advisable to remove compounds that might contaminate the isoprene product such as volatile organic compounds, acid gases, methane, benzene, toluene, ethylbenzene, xylenes, $H_2S$, COS, $CS_2$, HCl, $O_3$, organosulfur compounds, ammonia, nitrogen oxides, nitrogen-containing organic compounds, and heavy metal vapors. Removal of impurities from syngas can be achieved by one of several means, including gas scrubbing, treatment with solid-phase adsorbents, and purification using gas-permeable membranes.

### 3. Exemplary Cell Culture Conditions

**[0271]** Materials and methods suitable for the maintenance and growth of the recombinant cells of the invention are described *infra*, *e.g.*, in the Examples section. Other materials and methods suitable for the maintenance and growth of bacterial or cell cultures are well known in the art. Exemplary techniques can be found in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716, *Manual of Methods for General Bacteriology* Gerhardt *et al.*, *eds*), American Society for Microbiology, Washington, D.C. (1994) or Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc., Sunderland, MA. In some aspects, the cells are cultured in a culture medium under conditions permitting the expression of one or more MVA pathway, isoprene synthase, DXP pathway (*e.g.*, DXS), IDI, MVA pathway, polypeptides encoded by a nucleic acid inserted into the host cells.

**[0272]** Standard cell culture conditions can be used to culture the cells (*see*, for example, WO 2004/033646 and references cited therein). In some aspects, cells are grown and maintained at an appropriate temperature, gas mixture, and pH (such as at about 20°C to about 37°C, at about 6% to about 84% $CO_2$, and at a pH between about 5 to about 9). In some aspects, cells are

grown at 35°C in an appropriate cell medium. In some aspects, the pH ranges for fermentation are between about pH 5.0 to about pH 9.0 (such as about pH 6.0 to about pH 8.0 or about 6.5 to about 7.0). Cells can be grown under aerobic, anoxic, or anaerobic conditions based on the requirements of the host cells. In addition, more specific cell culture conditions can be used to culture the cells. For example, in some embodiments, the bacterial cells (such as *E. coli* cells) express one or more heterologous nucleic acids encoding mvaE and mvaS polypeptides under the control of a strong promoter in a low to medium copy plasmid and are cultured at 34°C.

**[0273]** Standard culture conditions and modes of fermentation, such as batch, fed-batch, or continuous fermentation that can be used are described in International Publication No. WO 2009/076676, U.S. Patent Application No. 12/335,071 (U.S. Publ. No. 2009/0203102), WO 2010/003007, US Publ. No. 2010/0048964, WO 2009/132220, US Publ. No. 2010/0003716. Batch and Fed-Batch fermentations are common and well known in the art and examples can be found in Brock, Biotechnology: *A Textbook of Industrial Microbiology*, Second Edition (1989) Sinauer Associates, Inc.

**[0274]** In some aspects, the cells are cultured under limited glucose conditions. By "limited glucose conditions" is meant that the amount of glucose that is added is less than or about 105% (such as about 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, or 10%) of the amount of glucose that is consumed by the cells. In particular aspects, the amount of glucose that is added to the culture medium is approximately the same as the amount of glucose that is consumed by the cells during a specific period of time. In some aspects, the rate of cell growth is controlled by limiting the amount of added glucose such that the cells grow at the rate that can be supported by the amount of glucose in the cell medium. In some aspects, glucose does not accumulate during the time the cells are cultured. In various aspects, the cells are cultured under limited glucose conditions for greater than or about 1, 2, 3, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, or 70 hours. In various aspects, the cells are cultured under limited glucose conditions for greater than or about 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 95, or 100% of the total length of time the cells are cultured. While not intending to be bound by any particular theory, it is believed that limited glucose conditions can allow more favorable regulation of the cells.

**[0275]** In some aspects, the bacterial cells are grown in batch culture. The bacterial cells can also be grown in fed-batch culture or in continuous culture. Additionally, the bacterial cells can be cultured in minimal medium, including, but not limited to, any of the minimal media

described above. The minimal medium can be further supplemented with 1.0 % (w/v) glucose, or any other six carbon sugar, or less. Specifically, the minimal medium can be supplemented with 1% (w/v), 0.9% (w/v), 0.8% (w/v), 0.7% (w/v), 0.6% (w/v), 0.5% (w/v), 0.4% (w/v), 0.3% (w/v), 0.2% (w/v), or 0.1% (w/v) glucose. Additionally, the minimal medium can be supplemented 0.1% (w/v) or less yeast extract. Specifically, the minimal medium can be supplemented with 0.1% (w/v), 0.09% (w/v), 0.08% (w/v), 0.07% (w/v), 0.06% (w/v), 0.05% (w/v), 0.04% (w/v), 0.03% (w/v), 0.02% (w/v), or 0.01% (w/v) yeast extract. Alternatively, the minimal medium can be supplemented with 1% (w/v), 0.9% (w/v), 0.8% (w/v), 0.7% (w/v), 0.6% (w/v), 0.5% (w/v), 0.4% (w/v), 0.3% (w/v), 0.2% (w/v), or 0.1% (w/v) glucose and with 0.1% (w/v), 0.09% (w/v), 0.08% (w/v), 0.07% (w/v), 0.06% (w/v), 0.05% (w/v), 0.04% (w/v), 0.03% (w/v), 0.02% (w/v), or 0.01% (w/v) yeast extract.

## 4. Exemplary Purification Methods

**[0276]** In some embodiments, any of the methods described herein further include a step of recovering the isoprene. For example, the isoprene produced using the compositions and methods of the invention can be recovered using standard techniques, such as gas stripping, membrane enhanced separation, fractionation, adsorption/desorption, pervaporation, thermal or vacuum desorption of isoprene from a solid phase, or extraction of isoprene immobilized or absorbed to a solid phase with a solvent (*see*, for example, U.S. Patent Nos. 4,703,007 and 4,570,029, which are each hereby incorporated by reference in their entireties, particularly with respect to isoprene recovery and purification methods). In one aspect, the isoprene is recovered by absorption stripping (*see, e.g.*, US Pub. No. 2011/0178261). In particular aspects, extractive distillation with an alcohol (such as ethanol, methanol, propanol, or a combination thereof) is used to recover the isoprene. In some embodiments, the recovery of isoprene involves the isolation of isoprene in a liquid form (such as a neat solution of isoprene or a solution of isoprene in a solvent). Gas stripping involves the removal of isoprene vapor from the fermentation off-gas stream in a continuous manner. Such removal can be achieved in several different ways including, but not limited to, adsorption to a solid phase, partition into a liquid phase, or direct condensation (such as condensation due to exposure to a condensation coil or do to an increase in pressure). In some embodiments, membrane enrichment of a dilute isoprene vapor stream above the dew point of the vapor resulting in the condensation of liquid isoprene.

In some embodiments, the recovery is performed as described in U.S. Patent Application Publication No. 2011/0178261. In some aspects, the isoprene is compressed and condensed.

**[0277]** The recovery of isoprene may involve one step or multiple steps. In some embodiments, the removal of isoprene vapor from the fermentation off-gas and the conversion of isoprene to a liquid phase are performed simultaneously. For example, isoprene can be directly condensed from the off-gas stream to form a liquid. In some embodiments, the removal of isoprene vapor from the fermentation off-gas and the conversion of isoprene to a liquid phase are performed sequentially. For example, isoprene may be adsorbed to a solid phase and then extracted from the solid phase with a solvent. In one aspect, the isoprene is recovered by using absorption stripping as described in U.S. Appl. No. 12/969,440 (US Publ. No. 2011/0178261).

**[0278]** In some embodiments, any of the methods described herein further include purifying the isoprene. For example, the isoprene produced using the compositions and methods of the invention can be purified using standard techniques. Purification refers to a process through which isoprene is separated from one or more components that are present when the isoprene is produced. In some embodiments, the isoprene is obtained as a substantially pure liquid. Examples of purification methods include (i) distillation from a solution in a liquid extractant and (ii) chromatography. As used herein, "purified isoprene" means isoprene that has been separated from one or more components that are present when the isoprene is produced. *See, e.g.* U.S. Patent Application Publication No. 2009/0203102, PCT publication WO 2009/076676 and U.S. Patent Application Serial No. 12/496,573. In some embodiments, the isoprene is at least about 20%, by weight, free from other components that are present when the isoprene is produced. In various embodiments, the isoprene is at least or about 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, 95%, or 99%, by weight, pure. Purity can be assayed by any appropriate method, *e.g.*, by column chromatography, HPLC analysis, or GC-MS analysis. In some aspects, at least a portion of the gas phase remaining after one or more recovery steps for the removal of isoprene is recycled by introducing the gas phase into a cell culture system (such as a fermentor) for the production of isoprene.

**[0279]** In some aspects, any of the methods described herein further include a step of recovering isoprene produced by any of the recombinant cells disclosed herein. In some aspects, the isoprene is recovered by absorption stripping (*See, e.g.,* U.S. Patent Application Publication No. 2011/0178261 A1).

**[0280]** In some aspects, any of the methods described herein further include a step of recovering the heterologous nucleic acid. In some aspects, any of the methods described herein further include a step of recovering the heterologous polypeptide.

**[0281]** Suitable purification methods are described in more detail in U.S. Patent Application Publication No. US2010/0196977 A1.

**[0282]** All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. It is also to be understood that one, some, or all of the properties of the various embodiments described herein may be combined to form other embodiments of the present invention. These and other aspects of the invention will become apparent to one of skill in the art.

**[0283]** Throughout this specification, various patents, patent applications and other types of publications (*e.g.*, journal articles) are referenced. The disclosure of all patents, patent applications, and publications cited herein are hereby incorporated by reference in their entirety for all purposes.

**[0284]** The invention can be more fully further understood by reference to the following examples, which are provided by way of illustration and are not meant to be limiting of the scope of the invention. All citations throughout the disclosure are hereby expressly incorporated by reference.

## EXAMPLES

**[0285]** The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

**Example 1: Identification of candidate isoprene synthases**

**[0286]** Amino acid sequences encoding isoprene synthase (IspS) isolated from *P. alba* and *P. montana* were analyzed against proprietary and public sequence databases to identify polypeptides that may have isoprene synthase activity. Screening criteria for identifying candidate polypeptides with potential isoprene synthase activity included: 1) the presence of one or more immutable amino acid residue(s) that correspond to the MEA *P. alba* isoprene synthase amino acid residues F287, G397, N438, E451, Y514; 2) sequence alignment based on mature sequences defined as polypeptides that had completed processing of immature signal sequences; and 3) sequence alignment based on the C-terminal region that contains the catalytically active site for IspS. Sequence alignments resulted in the identification of several candidate isoprene synthases including *A. hypogaea* (peanut), *G. max 1* (soybean), *G. max 2* (soybean), *Q. petraea* (oak), *Q. ilex, M. pruriens* (velvet bean), and *C. cajans* (pigeon pea) (Fig. 1, 2A-F and 3A-C).

**Example 2: Analysis of candidate isoprene synthases by assessment of isoprene synthase activity as measured by isoprene production in vitro.**

**[0287]** The candidate isoprene synthases were analyzed for the ability to convert DMAPP to isoprene *in vitro*.

*Materials and Methods*

**[0288]** Codon Optimization of IspS Enzymes and Strain Construction

**[0289]** DNA sequences encoding putative IspS enzymes from *A. hypogaea, C. cajans, G. max, M. pruriens* and *Q. petraea* were codon optimized, synthesized and cloned into the pCL201 expression vector by DNA2.0, in the identical orientation as the *P. alba* IspS MEA enzyme (Fig. 12). Purified plasmids were transformed into the expression host BL21 DE3 pLysS (Invitrogen) according to the manufacturer's recommended protocol, and resistant colonies harboring expression cassettes of the putative IspS enzymes were selected for further study. See Table 1 for detailed strain description.

Table 1. Strains used in this study

| Strain | Relevant Genotype |
|---|---|
| DW331 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-P. alba IspS MEA, Kan50, Chlor25 |
| DW668 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-A. hypogaea putative IspS, Kan50, |

| | |
|---|---|
| | Chlor25 |
| DW669 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-Q. petraea putative IspS, Kan50, Chlor25 |
| DW688 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-G. max 18280 putative IspS, Kan50, Chlor25 |
| DW689 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-G. max 21900 putative IspS, Kan50, Chlor25 |
| DW731 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-M. pruriens putative IspS, Kan50, Chlor25 |
| DW732 | BL21 DE3 pLysS (Invitrogen), pET24a PT7-C. cajan putative IspS, Kan50, Chlor25 |

**Isoprene Synthase Specific Productivity Determination**

*Materials*

**[0290]** Tris/ NaCl pH 7.6, MgCl2, 4-(2-Aminoethyl) benzenesulfonyl fluoride Hydrochloride (AEBSF), DNase I, DMAPP Triammonium salt(Cayman chemicals), Lysozyme (Sigma-Aldrich), 96-well Zinsser Glass Block, Seal & Sample Aluminum foil lids (Part No:538619)(Beckman coulter), Nunc MicroWell 96-Well Plates, Polypropylene, High Volume (Part No:2449946), (Thermo Scientific) (VWR).

**Protein expression and solubility measurement**

**[0291]** *E.coli* transformants expressing *P. alba* MEA isoprene synthase, *P. alba* variant 3 isoprene synthase, *A. hypogaea* isoprene synthase, *G. max 1* (G,max 21900 or Glyma09g21900.1) isoprene synthase, *G. max 2* (G,max 18280 or Glyma20g18280_1_FG) isoprene synthase, *M. pruriens* isoprene synthase, or *C. cajans* isoprene synthase were grown in LB media, induced at $OD_{600}$ ~ 0.5 with 200μM IPTG, and induced for 5 hours. Prior to harvesting the cells, a final $OD_{600}$ value was recorded. Cell pellets were collected by centrifugation and stored at -80°C. 3mL of lysis buffer (100 mM Tris, 100 mM NaCl pH 7.6, 1 mg/ml BSA, 1 mg/ml lysozyme, 0.1 mg/ml DNAase, 0.5 mM AEBSF, 5 mM MgCl2) was added to each frozen pellet, the pellets were re-suspended, and allowed to incubate on ice for 30 min. The cell suspension was then lysed fully by passing 2-3 times through a french pressure cell (small french press cell at 800 psi/Low setting) until lysate started to look clear. The mixture was then centrifuged at 14000 rpm for 25 minutes at 4°C. The supernatant was collected for use in

enzymatic activity assays. Samples from total lysate, the soluble fraction, and the pellet were analyzed for protein solubility by SDS-PAGE and Coomasie blue staining.

**Isoprene synthase enzymatic activity assay**

**[0292]** 25 µL of *E.coli* supernatant, containing isoprene synthase, was incubated with 0.25, 0.5, 1, 3, 5, 7, 10, 15, 20, and 25 mM DMAPP, in 100 µL reactions containing 50 mM $MgCl_2$ and 100 mM Tris/NaCl in a Zinsser 96-well glass block sealed with aluminum foil lids for 30 minutes at 34°C. The glass blocks were then transferred to an 80°C water bath for 2 minutes. Next, the glass blocks were analyzed by GC-FID (see below) to determine the concentration of isoprene generated in the reactions.

**[0293]** GC-FID Analysis

*Equipment and Materials*

**[0294]** Gas chromatograph (GC), 7890 (Agilent Technologies), Flame ionization detector (FID) 7890 (Agilent Technologies), HP-5ms column, 5%-phenyl-methylpolysiloxane, 15 m x 0.25 mm x 0.25 µm (Agilent Technologies), CTC autosampler (Leap Technologies), 0.2% v/v isoprene, balance nitrogen (Air Liquide), Chemstation with Enhanced Data Analysis (D.03.00.611)

*Procedure*

**[0295]** 96-well glass blocks were analyzed using GC-FID with the following parameters:

*Oven:*

| Rate (°C/min) | Temperature (°C) | Time (min) |
|---|---|---|
| 0 | 37 | 28 |

**[0296]** Run Time: 28 minutes

| FRONT INLET | |
|---|---|
| Front Inlet Temperature | 110°C |
| Flow Rate | 3.4 mL/min |
| Flow Mode | Constant Flow |
| Split Ratio | 50:1 |

| | |
|---|---|
| Carrier Gas | Helium |
| FLAME IONIZATION DETECTOR | |
| Detector Temperature | 160°C |
| Hydrogen Flow | 40 mL/min |
| Air Flow | 400 mL/min |
| Makeup Flow | 0.1 mL/min |
| Makeup Gas Type | Helium |

*Calculations*

**[0297]** The peak areas were converted to isoprene concentrations by dividing the peak area by the response factor calculated from 0.2% v/v isoprene in nitrogen calibration standards. Specific productivity of isoprene synthase was calculated in units of mg isoprene/L/hr/OD.

**Isoprene Synthase Specific Activity Determination**

**[0298]** Expression of 6XHis-tagged Isoprene Synthase

**[0299]** N-terminally 6XHis-tagged *P. alba* WT isoprene synthase, *P. alba* variant 1 isoprene synthase, *P. alba* variant 2 isoprene synthase, and *A. hypogaea* isoprene synthase were expressed and purified. The growth procedure is suitable for histidine tagged enzymes expressed in BL21(λDE3)pLysS cells. 10 ml of overnight culture was prepared for each 1 L of planned growth. The appropriate antibiotics (50 mg/ml kanamycin, 25 mg/ml chloramphenicol) were added to 10 ml of LB medium in a 25 ml flask and was inoculated with 1 colony from a fresh plate of cells or directly from glycerol frozen cell stock. Cultures were grown at 30 °C overnight with shaking at ~220 rpm. Day cultures were prepared in 1 liter of LB medium with appropriate antibiotics for each culture. Each 1L day culture was inoculated with 10 ml of overnight culture and grown at 30-37 °C with shaking at ~220 rpm until the OD600 reached ~0.4-0.6. Day cultures were then induced with 400 μM IPTG and allowed to continue growing at 30°C with shaking at 220 rpm for ~5-6 hours. Cells were then harvested by centrifugation at 10,000 x g for 10 min, 4 °C. Following harvest, cells were used directly or stored at -80°C until ready to process.

**Purification of 6XHis-tagged Isoprene Synthase**

**[0300]** For purification of histidine tagged enzymes from BL21(λDE3)pLysS cells, cells were gently resuspended in fresh Lysis buffer (Lysis buffer: Ni wash buffer + 0.5 mM PMST, 1 mg/ml lysozyme, 0.2 mg/ml DNaseI; Ni wash buffer: 50 mM NaH2PO4, 300 mM NaCl, 20 mM Imidazole, pH 8.0). Approximately 40-50 ml of lysis buffer was used per 1L of cell pellet. Cells were then incubated on ice for approximately 30 min. The cell suspension was then lysed fully by passing 2-3 times through a french pressure cell (large french press cell at 1200 psi/High setting) until lysate started to look clear. A sample of the lysate was saved for activity assay and gel analysis (~100 µl). The lysate was then clarified by centrifuging the lysate at 30,000xg for 30 min, 4°C in a Sorvall Discovery 90SE ultracentrifuge. The supernatant was removed and retained. A sample of the "clarified lysate" was saved for activity assay and gel analysis (~100 µl).

**[0301]** The clarified lysate was run over HisTrap HP columns (GE healthcare) using a gradient from 0-100% Ni buffer B. Following loading of the lysate on the column, the column was washed with Ni wash buffer (50 mM NaH2PO4, 300 mm NaCl, 20 mM imidazole, ph 8.0). The his-tagged Isoprene Synthase was then eluted from the column using a gradient from 0-100% Ni elution buffer (50 mM NaH2PO4, 300 mM NaCl, 500 mM imidazole, ph 8.0) and fractions containing the his-tagged Isoprene Synthase were collected. The column was then washed with Ni stripping buffer (20 mM NaH2PO4, 0.5 m NaCl, 50 mM EDTA, ph 7.4). Samples were then analyzed by SDS-PAGE gel (4-12% gel NUPAGE, Invitrogen) according to manufacturer's directions. Desired fractions were concentrated on spin filters (Vivaspin-20, Sartoris,) and then desalted over a Hi Prep 26/10 Desalting column (GE heathcare) packed with Sephadex G25 resin. The G-25 buffer consisted of 50 mM HEPES, 50 mM NaCl, and 1 mM DTT, pH 7.4. Fractions were then analyzed and concentrated. The samples were then stored at -80 °C.

**TEV Cleavage to Remove Histidine Tag**

**[0302]** Digestion was performed with TurboTEV Protease from Eton Bioscience Inc. One unit of TurboTEV per 10 µg of purified protein was used. The digest was performed at 4°C overnight. Samples were passed through another Ni column equilibrated in the Ni buffer to remove uncleaved enzyme, tag, TurboTEV protease (which is also tagged), and impurities. The

Ni column pass through and washes were analyzed using SDS-PAGE gel (NUPAGE, Invitrogen) and DMAPP activity assays. Samples containing pure enzyme were pooled and desalted into 50 mM NaCl pH 7.4 buffer, protein concentration was determined using the absorbance at $A_{280}$, and the purified protein was then aliquoted and stored at -80°C.

**Isoprene Synthase Enzymatic Activity Assay**

**[0303]** Isoprene Synthase enzymatic assay was performed as described above.

**Calculation of Isoprene Synthase Kinetic Parameters**

**[0304]** Isoprene synthase activity is reported in units of $s^{-1}$. Isoprene concentration is determined in the DMAPP assay and multiplied by a conversion factor of $1.47E^{-04}$ to obtain units of μM/L/second. Isoprene synthase activity is then calculated for each variant at each substrate concentration by:

$$Isoprene\ Synthase\ Activity = \frac{enzyme\ (\frac{\mu M}{L})}{isoprene\ (\frac{\mu M}{L/sec})}$$

**[0305]** Data from the isoprene synthase kinetic assays were fit to the following modified version of the Henri-Michaelis-Menten equation that takes into account substrate inhibition using Kaleidagraph 4.0 (Synergy Software) to determine $K_M$, $k_{cat}$ and $k_{iDMAPP}$ values for each isoprene synthase analyzed:

$$\frac{rate}{[Isoprene\ synthase]} = \frac{kcat * [DMAPP]}{KM + [DMAPP](1 + \frac{[DMAPP]}{KiDMAPP})}$$

*Results*

**[0306]** SDS-PAGE analysis of lysates prepared from cells expressing the indicated IspS demonstrated that *A. hypogaea* IspS (Fig. 4B; lanes 3 and 4) and *G. max 1* IspS (Fig. 4B; lanes 6 and 7) were soluble and expressed at levels comparable to *P. alba* MEA IspS (Fig. 4A; lane 2). In contrast, *G. max 2* IspS (Fig. 4B; lanes 9 and 10) was mostly insoluble as compared to *P. alba* MEA IspS. Analysis of three dilutions prepared from the soluble fraction of cells expressing *M.*

*pruriens* IspS (Fig. 5; lanes 5, 6 and 7) or *C. cajans* IspS (Fig. 5; lanes 8, 9 and 10)demonstrated that they were soluble and expressed at levels comparable to *P. alba* variant 3 IspS (Fig. 5; lanes 2, 3 and 4).

**[0307]** Production of isoprene from DMAPP that was provided to cell lysates containing the indicated IspS was measured. Lysates from cells expressing *A. hypogaea* IspS or *Q. petraea* IspS produced isoprene in the presence of increasing concentrations of DMAPP, with lysates containing *A. hypogaea* IspS demonstrating significantly higher production of isoprene as compared to lysates from cells expressing *Q. petraea* IspS (Fig. 6). Isoprene was produced from cell lysates containing *M. pruriens* IspS when in the presence of increasing concentrations of DMAPP as indicated by comparison of isoprene production by cells lysates containing the positive control, *P. alba* variant 3 IspS (Fig. 7). These results indicate that the candidate polypeptides, *A. hypogaea* IspS, *Q. petraea* IspS, and *M. pruriens* IspS have isoprene synthase activity.

**[0308]** Analysis of IspS specific productivity in lysates prepared from cells expressing the IspS candidates *A. hypogaea* (peanut), *G. max 1* (soybean), *G. max 2* (soybean), *M. pruriens* (velvet bean), or *C. cajans* (pigeon pea) demonstrated that *A. hypogaea* and *M. pruriens* had higher specific productivity than the other IspS candidates and *P. alba* IspS (Fig. 8 and Table 2). Cell lysates containing *G. max 1* IspS, *G. max 2* IspS, and *C. cajans* IspS did not produce isoprene in the presence of DMAPP.

Table 2. Comparison of IspS specific productivity

| | P. alba WT | | A. hypogaea | | G. max 1 | | G. max 2 | | M. pruriens | | C. cajans | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DMAPP (mM) | SP | Std Dev | SP | Std Dev | SP | Std Dev | SP | Std Dev | SP | Std Dev | SP | Std Dev |
| **0.25** | 3.7 | 0.173 | 1.2 | 0.132 | 0.000 | 0.000 | 0.000 | 0.000 | 1.1 | 0.013 | 0.000 | 0.000 |
| **0.5** | 5.6 | 0.167 | 2.1 | 0.027 | 0.000 | 0.000 | 0.000 | 0.000 | 1.8 | 0.045 | 0.000 | 0.000 |
| **1** | 8.3 | 0.340 | 4.0 | 0.164 | 0.000 | 0.000 | 0.000 | 0.000 | 4.0 | 0.160 | 0.000 | 0.000 |
| **3** | 13.2 | 0.782 | 11.0 | 0.378 | 0.000 | 0.000 | 0.000 | 0.000 | 10.1 | 0.710 | 0.000 | 0.000 |
| **5** | 14.4 | 0.860 | 14.6 | 3.445 | 0.201 | 0.014 | 0.171 | 0.004 | 15.3 | 0.328 | 0.202 | 0.006 |
| **7** | 14.6 | 0.437 | 22.4 | 0.561 | 0.272 | 0.030 | 0.244 | 0.025 | 17.5 | 0.582 | 0.291 | 0.031 |
| **10** | 13.4 | 0.761 | 26.9 | 0.660 | 0.397 | 0.032 | 0.322 | 0.018 | 19.8 | 2.569 | 0.427 | 0.004 |
| **15** | 11.1 | 0.299 | 30.8 | 1.013 | 0.591 | 0.007 | 0.473 | 0.021 | 19.0 | 3.933 | 0.570 | 0.020 |
| **20** | 9.4 | 0.358 | 33.1 | 0.733 | 0.802 | 0.065 | 0.598 | 0.003 | 19.0 | 1.648 | 0.765 | 0.014 |
| **25** | 7.8 | 0.217 | 32.6 | 0.596 | 0.958 | 0.075 | 0.784 | 0.034 | 17.3 | 0.274 | 0.928 | 0.008 |

SP = Specific Productivity (mg/L/h/OD)

**[0309]** Analysis of isoprene production when normalized to substrate provided to lysates prepared from cells expressing *P. alba* IspS demonstrated that at a concentration of about 6 mM DMAPP the enzyme exhibited substrate inhibition (Fig. 9). In contrast, *A. hypogaea* IspS did

not demonstrate substrate inhibition at the highest concentration of DMAPP assayed (Fig. 9). The lack of substrate inhibition by *A. hypogaea* IspS was dependent on the presence of magnesium ($Mg^{2+}$), a cofactor required by IspS for enzymatic activity. When *A. hypogaea* IspS was provided with 10 mM $Mg^{2+}$, substrate inhibition was observed at a concentration of about 10 mM DMAPP (Fig. 10). However, when 100 mM $Mg^{2+}$ was provided, no substrate inhibition was observed at the highest concentration of DMAPP assayed (Fig. 10).

**[0310]** Kinetic parameters of *A. hypogaea* IspS was determined *in vitro* with purified protein and compared to *P. alba* IspS (WT), *P. alba* variant 1 IspS, and *P. alba* variant 2 IspS. Analysis of isoprene production when normalized to substrate demonstrated that at a concentration of about 10 mM DMAPP, *P. alba* IspS exhibited substrate inhibition (Fig. 11 and Table 3). *P. alba* variant 1 IspS and *P. alba* variant 2 IspS both exhibited substrate inhibition at a concentration of about 15 mM DMAPP (Fig. 11 and Table 3). In comparison, *A. hypogaea* IspS exhibited substrate inhibition at a concentration of about 25 mM DMAPP (Fig. 11 and Table 3) with a higher $K_{iDMAPP}$ value at 30.6 than *P. alba* IspS (WT), *P. alba* variant 1 IspS, or *P. alba* variant 2 IspS, each having a $K_{iDMAPP}$ value of 2.17, 1.68, and 1.62, respectively (Table 4). In addition, *A. hypogaea* IspS had a $K_m$ value of 18.3 that was significantly higher than the $K_m$ value of *P. alba* IspS (WT), *P. alba* variant 1 IspS, or *P. alba* variant 2 IspS and the $K_{cat}$ value for *A. hypogaea* IspS was measured at 1.81 (Table 4).

Table 3. Comparison of IspS activity

| | P. alba WT | | P. alba variant 1 | | P. alba variant 2 | | A. hypogaea | |
|---|---|---|---|---|---|---|---|---|
| DMAPP (mM) | Activity (s-1) | Std Dev | Activity (s-1) | Std Dev | Activity (s-1) | Std Dev | Activity (s-1) | Std Dev |
| 0.25 | 0.208 | 0.009 | 0.166 | 0.009 | 0.168 | 0.006 | 0.018 | 0.001 |
| 0.5 | 0.366 | 0.011 | 0.295 | 0.023 | 0.277 | 0.003 | 0.036 | 0.001 |
| 1 | 0.560 | 0.013 | 0.427 | 0.021 | 0.407 | 0.025 | 0.077 | 0.001 |
| 3 | 0.951 | 0.025 | 0.751 | 0.018 | 0.734 | 0.024 | 0.233 | 0.012 |
| 5 | 1.079 | 0.028 | 0.871 | 0.037 | 0.878 | 0.045 | 0.371 | 0.016 |
| 7 | 1.144 | 0.048 | 0.892 | 0.048 | 0.930 | 0.030 | 0.468 | 0.014 |
| 10 | 1.128 | 0.047 | 0.911 | 0.020 | 0.962 | 0.034 | 0.584 | 0.007 |
| 15 | 0.965 | 0.014 | 0.779 | 0.030 | 0.878 | 0.043 | 0.677 | 0.016 |
| 20 | 0.833 | 0.008 | 0.667 | 0.023 | 0.783 | 0.009 | 0.695 | 0.007 |
| 25 | 0.646 | 0.025 | 0.558 | 0.019 | 0.674 | 0.031 | 0.680 | 0.008 |

Table 4. Comparison of IspS kinetic parameters

| | P. alba WT | | P. alba variant 1 | | P. alba variant 2 | | A. hypogaea | |
|---|---|---|---|---|---|---|---|---|
| | Avg Value | Avg Error | Avg Value | Avg Error | Avg Value | Avg Error | Avg Value | Avg Error |
| $K_{cat}$ | 2.17 | 0.309 | 1.68 | 0.192 | 1.62 | 0.166 | 1.81 | 0.502 |
| $K_m$ | 2.83 | 0.727 | 2.79 | 0.587 | 2.88 | 0.568 | 18.3 | 6.538 |
| $K_i$ | 13.5 | 3.5 | 14.9 | 3.3 | 22.1 | 5.1 | 30.6 | 17.1 |

**Example 3: Analysis of candidate isoprene synthases by DMAPP toxicity relief**

**[0311]** There is a strong correlation between increased intracellular DMAPP levels and growth inhibition of *E. coli*, which can be alleviated by the expression of *P. alba* IspS. Without being bound by theory, increased levels of IspS activity should therefore allow for better growth due to more rapid conversion of DMAPP to isoprene. The growth rates of *E. coli* expressing IspS candidates were monitored to identify candidates that display the ability to convert DMAPP to isoprene within the cell.

*Materials and Methods*

**[0312]** Construction of plasmid pEWL1036

**[0313]** The *A. hypogaea* IspS gene was optimized for expression in *E. coli* and synthesized by DNA2.0. To generate an expression construct for use in production hosts, *A. hypogaea IspS* was amplified from the DNA2.0 expression vector with PfuUltra II Fusion DNA Polymerase from Agilent Technologies (Santa Clara, CA) according to the manufacturer's protocol with

primers EL1304 and EL1305 (Table 5). The pDW34 backbone was amplified with primers EL1306 and EL1307 (Table 5). Using the GENEART Seamless Cloning and Assembly Kit from Life Technologies (Carlsbad, CA), the two PCR products were purified (Qiagen) and recombined to form an *in vitro* circular product. The product was electroporated into strain CMP451 according to standard molecular biology protocols. Cells were recovered in non-selective liquid LB medium, plated onto LA + 50μg/μl carbenicillin + 5mM mevalonic acid plates, and incubated overnight at 37°C. The next day, transformants were screened by PCR to check for the presence of the *A. hypogaea* IspS-mMVK fragment with primers EL1005 and EL1310 (Table 5). Transformants containing the correct sized PCR products were grown overnight in LB + 50μg/μl carbenicillin, and plasmids were purified for sequencing (Qiagen). The plasmids were verified by sequencing (Quintara Biosciences) with primers EL1004, EL1005, EL1006, EL1238, EL1308, EL1309, EL1310 (Table 5). One plasmid, pEWL1036, was selected for further study (Fig. 15 and Table 6).

Table 5. Primer sequences

| Primer name | Primer sequence | SEQ ID NO : |
|---|---|---|
| EL1004 | ACAATTTCACACAGGAAACAGC | 37 |
| EL1005 | CCAGGCAAATTCTGTTTTATCAG | 38 |
| EL1006 | GACAGCTTATCATCGACTGCACG | 39 |
| EL1238 | CGAAAAGCACCCTTATGTGTCTG | 40 |
| EL1304 | GGAATAAACCATGAACACCCGTCGCAGCGC | 41 |
| EL1305 | CTTTATGCAGTTAGTTGATCGGAATCGGTTCG | 42 |
| EL1306 | GATCAACTAACTGCATAAAGGAGGTAAAAAAACATGG | 43 |
| EL1307 | GGGTGTTCATGGTTTATTCCTCCTTATTTAATCG | 44 |
| EL1308 | CTGACCTACAAATTCGAAGAGG | 45 |
| EL1309 | GGTGGCGTGAGATCGGTCTG | 46 |
| EL1310 | GCATCGTCCGTTCGAGCTGC | 47 |

Table 6. Plasmid encoding *A. hypogaea* isoprene synthase

| **Plasmid name** | **Antibiotic resistance** | **Description** |
|---|---|---|
| pEWL1036 | Carbenicillin | pTrc *A. hypogaea* IspS-mMVK |

**[0314]** Construction of strains EWL1043, EWL1047, EWL1049, and EWL1052

**[0315]** For construction of strain EWL1043, the parent strain CMP1133 was co-transformed with plasmids pEWL1036 and MCM82 by electroporation according to standard molecular biology protocols. Cells were recovered in non-selective liquid LB, incubated for 2 hours at 30°C with shaking, and selected on LA + 50μg/μl carbenicillin + 50μg/μl spectinomycin plates and incubated overnight at 37°C. An individual colony resistant to carbenicillin and spectinomycin was designated as strain EWL1043 and used for further study (Table 7).

**[0316]** For construction of strain EWL1047, the parent strain CMP1133 was co-transformed with plasmids pEWL1036 and pCHL276 by electroporation according to standard molecular biology protocols. Cells were recovered in non-selective liquid LB, incubated for 2 hours at 30°C with shaking, and selected on LA + 50μg/μl carbenicillin + 50μg/μl spectinomycin plates and incubated overnight at 37°C. An individual colony resistant to carbenicillin and spectinomycin was designated as strain EWL1047 and used for further study (Table 7).

**[0317]** For construction of strain EWL1049, the parent strain CMP1133 was co-transformed with plasmids pEWL1036 and pCHL277 by electroporation according to standard molecular biology protocols. Cells were recovered in non-selective liquid LB, incubated for 2 hours at 30°C with shaking, and selected on LA + 50μg/μl carbenicillin + 50μg/μl spectinomycin plates and incubated overnight at 37°C. An individual colony resistant to carbenicillin and spectinomycin was designated as strain EWL1049 and used for further study (Table 7).

**[0318]** For construction of strain EWL1052, the parent strain CMP1133 was co-transformed with plasmids pEWL1036 and MCM1225 by electroporation according to standard molecular biology protocols. Cells were recovered in non-selective liquid LB, incubated for 2 hours at 30°C with shaking, and selected on LA + 50μg/μl carbenicillin + 50μg/μl spectinomycin plates and incubated overnight at 37°C. An individual colony resistant to carbenicillin and spectinomycin was designated as strain EWL1052 and used for further study (Table 7).

Table 7. Strains used in this study

| Strain name | Antibiotic resistance | Description |
|---|---|---|

| | | |
|---|---|---|
| EWL1043 | Carbenicillin, Spectinomycin | BL21, pgl-, PL.2-mKKDyI, GI1.2-gltA, yhfS-PyddV-ispA, pTrc *A. hypogaea* IspS-mMVK, pCL Ptrc-*E. faecalis* Upper MVA |
| EWL1047 | Carbenicillin Spectinomycin | BL21, pgl-, PL.2-mKKDyI, GI1.2-gltA, yhfS-PyddV-ispA, pTrc *A. hypogaea* IspS-mMVK, pCL Ptrc- leaderless *E. faecalis* Upper MVA |
| EWL1049 | Carbenicillin, Spectinomycin | BL21, pgl-, PL.2-mKKDyI, GI1.2-gltA, yhfS-PyddV-ispA, pTrc *A. hypogaea* IspS-mMVK, pCL Ptrc- leaderless *E. casseliflavus* Upper MVA |
| EWL1052 | Carbenicillin Spectinomycin | BL21, pgl-, PL.2-mKKDyI, GI1.2-gltA, yhfS-PyddV-ispA, pTrc *A. hypogaea* IspS-mMVK, pCL Ptrc-leaderless *E. gallinarum* Upper MVA |

**Growth and Specific Productivity Assay of Isoprene Production Strains**

**[0319]** The growth and isoprene productivity assay of isoprene production strains was performed as follows: Overnight cultures were prepared from glycerol culture stocks in LB medium with appropriate antibiotics in 30 ml glass test tubes and grown at 34°C. The next day, cultures were diluted to an OD600 of 0.2 in TM3 medium with appropriate antibiotics, distributed into a 48-well block (VWR), sealed with a Breathe Easier membrane (Diversified Biotech), and grown at 34°C at 600 rpm in a Shel Lab shaking incubator. Cultures were induced with appropriate amounts of IPTG when OD600 was between 0.4 and 0.8. Mevalonic acid was added subsequently to selected cultures one hour after induction. Both OD and isoprene were then measured every hour for 4 to 5 hours. OD600 was measured in a 96-well polystyrene plate (VWR) on a plate reader (Molecular Devices) according to standard molecular biology procedures. Isoprene measurements were conducted according to standard procedures. Briefly, 100 ul of bacterial culture was removed from the 48-well block and redistributed into a 96-well glass block (Zinsser) for analysis. Glass blocks were sealed with foil membranes, incubated on a Thermomixer (Eppendorf) for 30 minutes at 450 rpm at 34°C, and cultures were heat-killed for 2 minutes in an 80°C water bath. Glass blocks were cooled prior to standard isoprene measurement by GC-MS. Specific productivity was calculated by dividing the grams of isoprene produced by volume, cell density, and time.

*Results*

**[0320]** Cell growth and isoprene production in the presence of increasing MVA concentrations was measured over time in cells expressing *A. hypogaea* IspS (strain EWL1043). For *P. alba* variant 1 IspS expressing cells, growth was inhibited at any concentration of MVA tested (Fig. 13; first panel). For *P. alba* variant 2 IspS expressing cells, growth was inhibited when MVA was supplied at concentrations higher than about 15mM (Fig. 13, second panel). For *P. alba* variant 3 IspS expressing cells, growth was partially inhibited when MVA was supplied at MVA concentrations of at least about 30 mM and full inhibited at concentration of about 45 to 75 mM MVA (Fig. 13; third panel). For *A. hypogaea* IspS expressing cells, growth was partially inhibited at MVA concentrations of about 60 to 75 mM but significant growth was seen at MVA concentrations of about 15-45 mM (Fig. 13; fourth panel). Levels of cell growth corresponded to levels of specific productivity for isoprene, with *A. hypogaea* IspS expressing cells producing significant levels of isoprene as compared to *P. alba* variant1 IspS, *P. alba* variant 2 IspS, and *P. alba* variant 3 IspS (Fig. 14).

**[0321]** Cell growth and isoprene production was measured over time in cells co-expressing an *A. hypogaea* IspS or *P. alba* variant 1 IspS, with the *E. faecalis* upper MVA pathway (EWL1043 and EWL1047), the *E.casseliflavus* upper MVA pathway (EWL1049) or the *E.gallinarum* upper MVA pathway (EWL1052). For *P. alba* variant 1 IspS expressing cells, growth was reduced when the *E.casseliflavus* or *E.gallinarum* upper MVA pathway was co-expressed as compared to co-expression with the *E. faecalis* upper MVA pathway (Fig. 16; left panel). For *A. hypogaea* IspS expressing cells, co-expression with the *E.casseliflavus* or *E.gallinarum* upper MVA pathway resulted in comparable growth to cells co-expressing the *E. faecalis* upper MVA pathway (Fig. 16; right panel). Levels of cell growth corresponded to levels of specific productivity for isoprene, with *A. hypogaea* IspS expressing cells producing significant levels of isoprene as compared to *P. alba* variant1 IspS (Fig. 17).

**Example 4: Isoprene production from *E. coli* expressing legume isoprene synthase, grown in fed-batch culture at the 15-L scale.**

*Materials and Methods*

**[0322]** Medium Recipe (per liter fermentation medium):

**[0323]** K2HPO4 7.5 g, MgSO4 * 7H2O 2 g, citric acid monohydrate 2 g, ferric ammonium citrate 0.3 g, yeast extract 0.5 g, 50% sulphuric acid 1.6 mL, 1000X Modified Trace Metal Solution 1 ml. All of the components are added together and dissolved in Di H2O. This solution is heat sterilized (123°C for 20 minutes). The pH was adjusted to 7.0 with ammonium hydroxide (28%) and q.s. to volume. Glucose 10 g, Vitamin Solution 8 mL, and antibiotics are added after sterilization and pH adjustment.

**[0324]** 1000X Modified Trace Metal Solution (per liter):

**[0325]** Citric Acids * H2O 40 g, MnSO4 * H2O 30 g, NaCl 10 g, FeSO4 * 7H2O 1 g, CoCl2 * 6H2O 1 g, ZnSO * 7H2O 1 g, CuSO4 * 5H2O 100 mg, H3BO3 100 mg, NaMoO4 * 2H2O 100 mg. Each component are dissolved one at a time in Di H2O, pH is adjusted to 3.0 with HCl/NaOH, and then the solution is q.s. to volume and filter sterilized with a 0.22 micron filter.

**[0326]** Vitamin Solution (per liter):

**[0327]** Thiamine hydrochloride 1.0 g, D-(+)-biotin 1.0 g, nicotinic acid 1.0 g, D-pantothenic acid 4.8 g, pyridoxine hydrochloride 4.0 g. Each component is dissolved one at a time in Di H2O, pH is adjusted to 3.0 with HCl/NaOH, and then the solution is q.s. to volume and filter sterilized with 0.22 micron filter.

**[0328]** Feed solution (per kilogram):

**[0329]** Glucose 0.57 kg, Di H2O 0.38 kg, K2HPO4 7.5 g, and 100% Foamblast 10 g. All components are mixed together and autoclaved. Macro Salt Solution 3.4 mL, 1000X Modified Trace Metal Solution 0.8 ml, and Vitamin Solution 6.7 mL are added after the solution had cooled to 25°C.

**[0330]** Macro Salt Solution (per liter):

**[0331]** MgSO4 * 7H2O 296 g, citric acid monohydrate 296 g, ferric ammonium citrate 49.6 g. All components are dissolved in water, q.s. to volume and filter sterilized with 0.22 micron filter.

**[0332]** Fermentation is performed in a 15-L bioreactor with *E. coli* BL21 cells overexpressing the upper MVA pathway (pCLUpper – MCM82), the lower MVA pathway (PL.2-mKKDyI) and isoprene synthase from *A. hypogaea* (pET24a PT7-A).

**[0333]** This experiment is carried out to monitor isoprene formation from glucose at the desired fermentation pH 7.0 and temperature 34°C. A frozen vial of the *E. coli* strain is thawed and inoculated into a flask with tryptone-yeast extract medium and the appropriate antibiotics. After the inoculum grows to optical density 1.0, measured at 550 nm ($OD_{550}$), 500 mL is used to inoculate a 15-L bioreactor and bring the initial tank volume to 5 L.

**[0334]** The feed solution is fed at an exponential rate until a top feed rate of 6 g/min is reached. After this time, the glucose feed is fed to meet metabolic demands at rates less than or equal to 6 g/min. The total amount of glucose delivered to the bioreactor during the 52 hr fermentation is 6.8 kg. Induction is achieved by adding isopropyl-beta-D-1-thiogalactopyranoside (IPTG). A shot of IPTG is added to the tank to bring the concentration to 100 uM when the cells are at an $OD_{550}$ of 6 and a second shot is added bring the concentration to 100 uM when the cells are at an $OD_{550}$ of 100.

**[0335]** The isoprene level in the off-gas from the bioreactor is determined using an iSCAN (*Hamilton* Sundstrand) mass spectrometer.

## SEQUENCES

*Amino Acid Sequence of MEA P. alba IspS*

MEARRSANYEPNSWDYDYLLSSDTDESIEVYKDKAKKLEAEVRREINNEKAEFLTLLEL
IDNVQRLGLGYRFESDIRGALDRFVSSGGFDAVTKTSLHGTALSFRLLRQHGFEVSQEAF
SGFKDQNGNFLENLKEDIKAILSLYEASFLALEGENILDEAKVFAISHLKELSEEKIGKEL
AEQVNHALELPLHRRTQRLEAVWSIEAYRKKEDANQVLLELAILDYNMIQSVYQRDLR
ETSRWWRRVGLATKLHFARDRLIESFYWAVGVAFEPQYSDCRNSVAKMFSFVTIIDDIY
DVYGTLDELELFTDAVERWDVNAINDLPDYMKLCFLALYNTINEIAYDNLKDKGENILP
YLTKAWADLCNAFLQEAKWLYNKSTPTFDDYFGNAWKSSSGPLQLVFAYFAVVQNIK
KEEIENLQKYHDTISRPSHIFRLCNDLASASAEIARGETANSVSCYMRTKGISEELATESV
MNLIDETWKKMNKEKLGGSLFAKPFVETAINLARQSHCTYHNGDAHTSPDELTRKRVL
SVITEPILPFER (SEQ ID NO:1)

*Nucleic Acid Sequence of MEA P. alba IspS*

atggaagcacgtcgctctgcgaactacgaacctaacagctgggactatgattacctgctgtcctccgacacggacgagtccatcgaagtat
acaaagacaaagcgaaaaagctggaagccgaagttcgtcgcgagattaataacgaaaaagcagaatttctgaccctgctggaactgattg
acaacgtccagcgcctgggcctgggttaccgtttcgagtctgatatccgtggtgcgctggatcgcttcgtttcctccggcggcttcgatgcgg
taaccaagacttccctgcacggtacggcactgtctttccgtctgctgcgtcaacacggttttgaggtttctcaggaagcgttcagcggcttcaa
agaccaaaacggcaacttcctggagaacctgaaggaagatatcaaagctatcctgagcctgtacgaggccagcttcctggctctggaagg
cgaaaacatcctggacgaggcgaaggttttcgcaatctctcatctgaaagaactgtctgaagaaaagatcggtaaagagctggcagaaca
ggtgaaccatgcactggaactgccactgcatcgccgtactcagcgtctggaagcagtatggtctatcgaggcctaccgtaaaaaggagga
cgcgaatcaggttctgctggagctggcaattctggattacaacatgatccagtctgtataccagcgtgatctgcgtgaaacgtcccgttggtg
gcgtcgtgtgggtctggcgaccaaactgcactttgctcgtgaccgcctgattgagagcttctactgggccgtgggtgtagcattcgaaccgc
aatactccgactgccgtaactccgtcgcaaaaatgttttctttcgtaaccattatcgacgatatctacgatgtatacggcaccctggacgaactg
gagctgtttactgatgcagttgagcgttgggacgtaaacgccatcaacgacctgccggattacatgaaactgtgctttctggctctgtataaca
ctattaacgaaatcgcctacgacaacctgaaagataaaggtgagaacatcctgccgtatctgaccaaagcctgggctgacctgtgcaacgc
tttcctgcaagaagccaagtggctgtacaacaaatctactccgacctttgacgactacttcggcaacgcatggaaatcctcttctggcccgct
gcaactggtgttcgcttacttcgctgtcgtgcagaacattaaaaaggaagagatcgaaaacctgcaaaaataccatgacaccatctctcgtcc
ttcccatatcttccgtctgtgcaatgacctggctagcgcgtctgcggaaattgcgcgtggtgaaaccgcaaatagcgtttcttgttacatgcgc
actaaaggtatctccgaagaactggctaccgaaagcgtgatgaatctgatcgatgaaacctggaaaaagatgaacaaggaaaaactgggt

ggtagcctgttcgcgaaaccgttcgtggaaaccgcgatcaacctggcacgtcaatctcactgcacttatcataacggcgacgcgcatacctc
tccggatgagctgacccgcaaacgcgttctgtctgtaatcactgaaccgattctgccgtttgaacgctaa (SEQ ID NO:2)

*Amino Acid Sequence of A. hypogaea IspS*

MNTRRSANYQPNLWDFEFLQSVENDLQVERLEERARKLEEEVRGLMKKVEIEPLSLLEL
MDNVERLGLTYKFEEDIKSALNNRIVPLLHHHTINKYGLHATALSFRFLRQHAFHVSPD
VFESFKEEGKFKKEISGDVLGLLNLYETSYLGFEGETILDEARAFSATHLKNLLQTNQVQ
NKVMAEKVRHALELPYHRRVHRLEARWFIERYEQKEAHDGALLELAKLDFNMVQSV
MKKELQELSRWWREIGLTSKLDFVRDRLMEVYFWALGMAPHPQLTECRKAVTKMFGL
VTIIDDVYDVYGTLDELQLFTDAVDRWDVNAVETLPDYMKLCYLALYNSVNDTAYST
LREKGDNSLPHLAKSWRDLCKAFLQEAKWSNNKIIPPFDAYIRNASVSSSGGALLAPCY
FSVTQDSTSQAIDSITNYHGIVRSSCAIFRLCNDLATSAAELERGETTNSITSYMTENGTT
EEEARESLGKLIDQEWKKMNRDVVLESAYPNVFKEIAINMARVSHCTYQYGDGLGRPD
DTAENRIKLSLIEPIPIN (SEQ ID NO:3)

*Nucleic Acid Sequence of A. hypogaea IspS*

atgaacacccgtcgcagcgccaattaccagccgaacctgtgggatttcgagtttttgcaaagcgtcgaaaacgacttgcaggtggaacgcc
tggaagagcgtgcacgtaagctggaagaagaagtgcgtggtctgatgaaaaaggtcgagattgagccgttgagcctgctggaactgatgg
acaacgttgagcgcctgggtctgacctacaaattcgaagaggacatcaaaagcgcgttgaataaccgcattgtcccactgttgcatcaccat
actatcaataagtacggtctgcacgccacggctctgagcttccgtttcctgcgtcaacacgcctttcacgtcagcccggatgtttttgaaagctt
caaagaagaaggtaagttcaagaaagagattagcggcgacgtgctgggtctgctgaacctgtacgagactagctacctgggctttgaaggt
gaaaccattctggacgaggcacgcgccttcagcgctacccatctgaaaaatctgttgcaaaccaaccaggtgcagaataaagttatggcgg
agaaggtccgccatgcgttggagctgccgtatcaccgtcgtgttcaccgtttggaagcccgctggtttattgagcgctatgagcagaaagag
gcgcatgatggtgccttgctggagctggcgaaactggatttcaacatggttcagtctgtgatgaagaaagagctgcaagagctgagccgct
ggtggcgtgagatcggtctgaccagcaagctggacttcgtgcgtgatcgtctgatggaagtgtacttttgggcgctgggtatggctccgcac
ccgcagctgacggagtgccgtaaagcagtgaccaagatgtttggcctggttaccatcattgacgatgtttacgatgtgtatggcaccctgga
cgagctgcaactgtttacggatgcggttgaccgttgggacgttaacgcagtcgaaacgctgccggactacatgaaactgtgttacctggcg
ctgtataactccgttaatgacacggcatatagcactctgcgtgagaagggtgacaatagcctgcctcacttggcaaagtcgtggcgtgatctg
tgtaaggcgtttctgcaagaggcgaagtggagcaataacaagatcattccgccgttcgatgcgtacatccgcaacgcatctgtcagcagca
gcggcggtgctctgttggcgccatgttacttctccgttacgcaagacagcacgagccaggccatcgattctattacgaactaccacggcatc
gtccgttcgagctgcgcaatcttccgcctgtgcaatgacctggcgacctctgctgcggagctggagcgtggcgaaaccaccaattccatca
cgtcctatatgaccgaaaatggcaccaccgaagaagaggcgcgtgaaagcctgggtaaactgattgaccaagagtggaagaaaatgaat

cgtgatgtggtcctggaaagcgcgtatccgaacgtgtttaaagaaattgcgattaacatggcacgcgttagccattgcacctatcagtatggc gatggtctgggtcgtccggatgatactgcggagaatcgtatcaagctgtctctgatcgaaccgattccgatcaactaa (SEQ ID NO:4)

*Amino Acid Sequence of G. max 1 IspS*

METRRSANYQPNLWNFEFLPPSLENDHKVEKLEERARKVEEEVRRMINGADTEALRLL ELIDEIQRLGLTYKFEKDIFKALEKTISLDENEKHISGLHATALSFRLLRQHGFEVSQDVF KRFKDKEGGFINELKGDMQGLLSLYEASYLGFEGETLLDEARAYSITHLKNNLKVGVN TEVKEQVSHALELPYHRGLNRLEARWFLEKYEPNESHHHVLLELAKIDFNLVQVMYQK ELRELSRWWSEMGLTSKLKFVRDRLMEVYFWVLGMAPRPQFSECRKAVTKTFALIGII DDVYDVYGTLDELQLFTDAIERWDVNAMNTLPDYMKLCYLAVYNTVNDTCYSTLKA KGHNNMSYLTKSWCELCKAFLQEAKWSNNKIVPTFSKYLENASVSSSGMALLTASYFS VCQQQDISNQQALCSLTNFQGLVRSSSNIFRLCNDLATSAAELETGETANSITCYMHEKD TSEEQAREELTNLIDAEWKKMNREFVSNSTLPKAFKEIAINMARVSHCMYQYEDGLGR PGYTTENKIKLLLIDPVPIN (SEQ ID NO:5)

*Nucleic Acid Sequence of G. max 1 IspS*

atggaaactcgtcgctctgccaattaccaaccgaacctgtggaactttgaatttttgccgccgtccctggagaatgaccacaaggttgagaaa ttggaagaacgtgcacgcaaggttgaagaagaagtccgtcgcatgatcaacggcgcagataccgaggctctgcgcctgctggaactgatt gatgagattcaacgtttgggtctgacctacaagtttgagaaggacatcttcaaggctctggagaaaaccattagcctggacgagaacgaga aacacattagcggtctgcacgcgacggcgctgtcctttcgcctgctgcgccagcacggcttcgaggtaagccaggacgtgttcaaacgctt taaggacaaagaaggcggttttatcaatgagctgaaaggtgatatgcagggtctgctgagcctgtacgaagcgtcgtatttgggctttgaag gtgaaacgctgctggatgaagcccgtgcgtatagcatcacccacctgaaaaacaacttgaaagtgggtgttaataccgaggtgaaagagc aggtgtcccacgcactggaactgccgtaccatcgcggcctgaatcgtttggaagctcgttggttcctggagaagtatgagccgaacgagag ccatcatcatgtgctgctggagctggcgaagatcgactttaacctggttcaggtcatgtaccagaaagaactgcgtgagttgagccgttggtg gagcgaaatgggcctgaccagcaagctgaagtttgtccgtgatcgcctgatggaagtctacttttgggtcttgggtatggcaccgcgtccgc agttcagcgagtgccgtaaggcggtgaccaaaaccttcgccctgattggtatcatcgacgatgtgtatgacgtctatggcaccctggatgag ctgcaattgttcaccgatgcgatcgagcgttgggacgttaatgcaatgaatacgctgccggactatatgaagctgtgttacttggcagtttaca ataccgtgaatgatacctgttactctaccctgaaggcaaaaggccacaataacatgtcctatctgaccaagagctggtgtgagctgtgcaaa gccttcctgcaagaggcgaagtggagcaacaacaaaattgttccgaccttcagcaaatacttggagaacgcgagcgtcagcagctcgggt atggccctgctgaccgcgagctatttctccgtgtgtcagcaacaggacatttctaatcagcaagcactgtgcagcctgacgaattttcaaggt ctggttcgttctagcagcaacattttccgtctgtgcaatgatctggcgacgtctgcggcggaactggaaactggtgagactgccaacagcat

cacctgctatatgcacgagaaagacacgagcgaagagcaagcgcgtgaagaattgacgaacctgatcgacgccgagtggaagaaaatg aaccgcgagttcgtgagcaattcgaccctgccaaaagctttcaaagagattgcgatcaatatggcgcgtgttagccattgcatgtaccaatac gaggatggcctgggtcgtccgggttacacgacggagaacaagattaagctgctgctgattgatccggtccctatcaattaa (SEQ ID NO:6)

*Amino Acid Sequence of G. max 2 IspS*

METRRSANYQPNLWNFEFLPPSLENDHKVEKLEERAKKVEEEVRKVINGIDTKPLLLELI DDVQHLGLTYKFEKDIIKALEKIVSLDENEEHKSELYYTALSFRLLRQHGFEVSQDVFKR FKDKEGGFSGELKGDVQGLLSLYEASYLGFEGDNLLDEARAFSTTHLKNNLKQGINTKE AEQVNHALELPYHRRLQRLEARWYLEKYEPKEPHHQLLLELAKLDFNMVQLLHQKEL QELSRWWSEMGLASKLEFARDRLMEVYFWALGMAPDPQFRECRKAVTKMFGLVTIID DVYDIYGTLDELQLFTDAVERWDVNVVNTLPDYMKLCYLALYNTVNDTAYSILKEKG RNNLSYLKKSWCELCKAFLQEAKWSNNKIVPAFSKYLENASVSSSGVALLAPSYFSVCQ EQDISFSDKTLHYLTNFGGLVRSSCTIFRLCNDLTTSAAELERGETTNSIMSYMHENGTS EEHACEELRNLIDIEWKKMNRQRVSDSTLPKAFREIAMNMARVSHNTYQYGDGLGRPD YNIENRIKFLLIDPVPIN (SEQ ID NO:7)

*Nucleic Acid Sequence of G. max 2 IspS*

atggaaacccgtcgtagcgccaattatcaacctaacctgtggaattttgagtttttgccgccgtccctggagaatgatcacaaggtggaaaaa ctggaagagcgcgcgaagaaggtcgaagaagaggtccgcaaggtcatcaatggcattgataccaaaccgctgttgctggagttgatcga cgatgtgcaacatctgggtctgacctataagtttgagaaggacatcattaaggcgctggagaagattgttagcctggatgagaacgaagag cacaaaagcgaattgtattacaccgcgctgagctttcgcctgctgcgtcagcatggctttgaagtgtcccaggacgttttcaaacgtttcaagg ataaagagggtggtttcagcggtgaactgaaaggcgacgtccaaggcttgctgtcgctgtatgaagcgtcgtacctgggtttcgagggcga taacctgctggacgaggcacgtgcattttctacgacgcacctgaagaacaatttgaagcagggtattaacaccaaagaggcggagcaagtt aaccacgcactggaactgccgtatcaccgtcgtctgcaacgtctggaagcgcgctggtatctggagaaatacgaaccaaaagagccgca ccaccaactgctgttggaactggctaaattggacttcaacatggtgcagctgctgcatcagaaagaattgcaggagctgtctcgttggtgga gcgagatgggtctggcaagcaagctggagtttgcgcgcgatcgcctgatggaagtgtacttttgggcactgggtatggcgcctgacccgc agttccgtgagtgtcgtaaggcggttaccaaaatgttcggtctggtcaccatcattgacgatgtttacgacatttacggtacgctggacgaact gcaactgtttacggacgccgtggagcgttgggacgtcaatgttgtgaacacgctgccggactatatgaaactgtgctacttggccttgtacaa cacggttaatgatactgcgtactctattctgaaagagaaaggccgcaacaatctgagctatttgaagaaaagctggtgcgaactgtgcaaag cctttctgcaagaggctaagtggtccaacaataagattgtgccggcattcagcaaatacctggaaaatgcaagcgtcagcagcagcggcgt cgctctgctggcgccgagctacttcagcgtgtgtcaggagcaggatattagcttcagcgacaagaccctgcactatctgacgaatttcggtg

gtctggtgcgctctagctgtaccattttccgtctgtgcaatgacctgacgaccagcgcagcggagctggaacgcggtgaaaccactaatag
cattatgtcctatatgcacgagaacggtaccagcgaagagcatgcctgcgaagagttgcgtaacctgatcgacatcgagtggaagaagatg
aaccgccaacgtgtttcggatagcaccctgccgaaggctttccgtgagatcgcgatgaacatggcccgtgtttctcataacacctaccagta
cggcgatggcctgggtcgtccggattacaacatcgagaatcgcatcaaatttctgctgatcgatccagttccgatcaat (SEQ ID
NO:8)

*Amino Acid Sequence of M.pruriens IspS*

MATKVLCLSNQFLYPTPTLTSTRFLQTENFTQKTSLINPKPYPLFCVVTSQFSQITEDNTR
RSANYHPNLWNFEFLQSLENDPKIEKLEEKATKLVEEVRHMMNKAETEPLSLLELIDDV
QRLGLTYKFEKDIINALEKTISLDENQKHISGLHATSLSFRLLRQHGFEVSQDVFKKFKD
EDGGFSAELKGDVQGLLSLYEASYLGFEGENLLDEAREFSIEHLKNNLNKGITTKVAEQ
VSHALELPYHRRIHRLEARWFLDKYEPKESQHKLLLELAKLDFNMVQSLHQKELRELS
MWWREIGLTSKLDFVRDRLMEVYFWALGMAPDPQFSECRKAVTKMFGLVTIIDDVYD
VYGTLDELQLFTDAVERWDVNAINTLPDYMKLCFLALYNTVNDTTYSILKEKGHNNIS
YLTKSWCELCKAFLQEAKWSNNKIIPTFNKYLRNASVSSSGVALLAPSFFLVCQEQDISE
QALHSLINFHGLVRSSCVIFRLCNDLATSAAELERGETTNSITSYMHENGTSEEQARQEL
RILIDAEWKNMNQERYLDSTLPDAFMEITINLARVSHCTYQYGDGLGRPDYTTKNRIKL
LLIDPLPIN (SEQ ID NO:9)

*Nucleic Acid Sequence of M.pruriens IspS*

atgaacacgcgtcgctcggccaactatcacccaaacctgtggaacttcgaattcttgcaaagcctggagaatgatccgaagatcgaaaagc
tggaagagaaggcgacgaagctggtcgaagaggttcgtcacatgatgaataaggcggaaaccgagccgctgagcctgctggaactgat
cgacgacgtgcagcgcctgggtttgacctacaagtttgaaaaagacatcattaatgcactggagaaaacgattagcctggatgagaaccaa
aagcacattagcggcttgcatgccacgagcctgtcttttcgtctgctgcgccaacacggttttgaggtgtctcaagatgtgttcaaaaagtttaa
agatgaggacggtggtttcagcgcggaactgaagggcgacgttcagggtctgctgagcttgtacgaggcgagctatctgggctttgaggg
tgagaatctgctggatgaagcgcgcgaattttccatcgaacacctgaaaaacaatctgaacaagggtattacgaccaaagtggcggaacaa
gtgagccacgccctggagctgccgtatcaccgccgcatccatcgcctggaagcgcgttggttcctggacaaatacgaaccgaaagagtcc
cagcataagctgctgttggagctggcgaaactggatttcaacatggtgcagagcctgcatcaaaaagagctgcgcgagctgagcatgtggt
ggcgtgagattggcctgacctctaagctggacttcgtccgtgatcgtttgatggaagtttacttttgggcactgggcatggcaccggacccgc
aattttctgaatgtcgtaaagcagtgactaaaatgttcggtttggtgaccatcattgatgacgtctacgatgtttatggtacgctggatgagctgc
aactgttcactgacgcggtcgagcgttgggacgtcaatgctatcaataccctgccggactatatgaagctgtgctttttggctctgtacaacac
ggtcaacgacactacctacagcatcctgaaagaaaagggtcacaataacatcagctacttgaccaaatcctggtgcgagctgtgcaaagct

tttctgcaagaagcgaagtggagcaataacaaaatcattccaaccttcaataagtatctgcgtaatgcgagcgtaagcagcagcggcgttgc
cctgctggcaccttctttcttcctggtctgccaggagcaggatatttccgagcaggcgttgcattccctgattaactttcacggtctggttcgca
gcagctgtgttatcttccgtttgtgcaatgatctggctacgagcgcagcggagctggagcgtggtgaaacgaccaacagcattaccagctat
atgcacgagaatggcaccagcgaagagcaggcacgtcaggaactgcgtattttgatcgacgcggagtggaaaaacatgaatcaggaac
gttatctggatagcacgctgccggatgccttcatggagattaccatcaacctggcccgtgtttcgcattgtacctaccagtacggcgacggtc
tgggccgtccggactacaccaccaagaaccgcattaaactgctgctgatcgacccgctgccgatcaattaa (SEQ ID NO:10)

*Amino Acid Sequence of C. cajans IspS*

MATHHLLCLSNPFSSPSPTLSTATRSFPLTNNFNHKTSLANSKPCPFICSQITHHHHTRRS
ANYQPNLWNFEFLQSLQNHHQVFTMFRRKLEKEVRCMMNKADAEALSLLELIDDVQR
LGLTYRFEKDIIKVLEKIVSLDEIEKHQSGLHATALTFRLLRQHGFHQVSQDMFKRFKDK
EGGFNDELKGDVQGLLSLYEASYLGFEGEYLLDEARAFSITHLNNSLKQGINTKLAEQV
SHALQLPHHRRLHRLEARWQLDKYEPKEPHHHLLLHLAKLDFNILQSLYQNELRELSR
WWREMGLTSKLEFVRDRLMEVYFWALGMAPHPEFSECRKAITKMFGLVTIIDDVYDV
YGTLDELQLFTDAVERWDVNVVNTLPYYMKLCYLALYNTVNETSYSILKENGHNSLSY
LAKSWCELCKAFLEEAKWSKKKVIPALNRYLENAWVSSSGVALLAPCYFSVCKEEDKI
SDEALHSLTNFHGLVRSSCAIFRLYNDLATSAAELERDETTNSMTCYMHENGSCEEQAR
EELRKMIEVEWKKMNQEGVLDCTLPTAFKEIAMNMARVSHCTYQHGDGLGRPDYTTQ
NRIKLLLIDPLPIN (SEQ ID NO:11)

*Nucleic Acid Sequence of C. cajans IspS*

atgcacactcgccgctcggccaactaccaaccaaacctgtggaacttcgaattcttgcaatccctgcaaaatcaccatcaggtgtttacgatg
tttcgtcgtaaactggagaaagaagtgcgttgcatgatgaataaagcggatgcggaagcgctgtctctgctggagttgattgatgacgtgca
gcgcctgggtctgacctaccgttttgagaaagatatcattaaggtcctggagaagatcgttagcctggatgagatcgagaaacaccagagc
ggtctgcacgccacggcattgaccttccgcctgttgcgccaacatggctttcaccaggttagccaggatatgttcaaacgtttcaaagacaaa
gagggtggtttcaatgatgaactgaagggcgatgttcagggtctgctgtccctgtacgaagcaagctatctgggcttcgagggcgaatacct
gctggacgaggcacgcgcgttcagcattacccacctgaacaacagcctgaaacagggcatcaacaccaagctggcagagcaggtgagc
catgctctgcaactgccgcatcaccgccgtctgcaccgtctggaagctcgttggcagctggataagtatgagccgaaagaaccgcaccatc
atctgctgctgcatctggcgaagctggactttaacattctgcaaagcctgtatcaaaacgagctgcgtgagttgagccgttggtggcgtgaga
tgggcctgaccagcaagctggagtttgttcgtgatcgcctgatggaagtctacttttgggctctgggtatggctccgcatccagagttcagcg
aatgccgcaaagcaattaccaaaatgtttggtttggtcaccattatcgacgacgtgtatgacgtctacggtaccctggacgagctgcaactgtt
tacggacgcggttgagcgttgggacgttaatgttgttaataccctgccttactatatgaagctgtgctatctggcattgtacaacaccgtgaatg

aaacgagctactctattctgaaagaaaatggtcacaatagcctgagctacttggcgaagagctggtgtgaactgtgcaaggcgttcctggaa gaggccaagtggagcaagaagaaagtcatcccggcgttgaaccgttatttggagaacgcgtgggttagctcgtccggcgtggcgctgctg gccccgtgctatttctccgtgtgtaaagaagaggacaaaatctccgacgaggctctgcatagcctgaccaactttcacggcctggtgcgtag cagctgcgccatcttccgtctgtacaacgatttggcgacgagcgcagcggagctggagcgcgacgaaacgacgaatagcatgacttgtta catgcacgaaaatggttcttgtgaagagcaggcgcgtgaagagctgcgcaagatgatcgaggtcgaatggaaaaagatgaatcaggaag gtgtcctggattgcaccctgccgacggccttcaaagagattgcgatgaacatggcacgtgtgagccactgtacctatcaacacggtgatggt ctgggccgtccggactacaccacgcaaaatcgtattaagctgctgttgatcgatccgctgccgattaattaa (SEQ ID NO:12)

*Amino Acid Sequence of Q. petraea IspS*

MTERQSANFQPSLWSYEYIQSLKNGYEADLYEDRAKKLGEEVRRMINNKDTKLLTTLE LIDDIERLGLGYRFKEEIMRALDRFVTLKGCEEFTNGSIHDTALSFRLLRQHGFGVSQDM FNCFKDQKGNFKECLSKDIKGLLSLYEASYLGFEGENLLDEAREFTTMHLKDLKGDVSR TLKEEVRHSLEMPLHRRMRRLEQRWYIDAYNMKEAHDRKLLELAKLDFNIVQSVHQR DLKDMSRWWQEMGLGNKLSFARDRLMECFFFSVGMAFEPQFSNSRKAVTKMFSFITVI DDIYDVYATLEELEMFTDIVQRWDVKAVKDLPEYMKLCFLALFNTVNEMVYDTLKEQ GVDILPYLTKAWGDICKAFLQETKWRYYKRTPSSEDYLDNAWISVSGALLLIHAYFLMS PSITDRALKGLEDYHNILRWPSIIFRLTNDLGTSTAELERGETANSILCYMRETSRSEDFA REHISNLIDKTWKKMNKDRFSDSPFEEPFLETAINLARISHCIYQHGDGHGAPDTRTKDR VLSLIIEPIPCYDPSTNFHSQIHL (SEQ ID NO:13)

*Nucleic Acid Sequence of Q. petraea IspS*

atgaccgaacgtcaaagcgcgaactttcaaccgtccctgtggagctacgaatacattcagagcctgaagaatggctatgaggcggacctgt atgaagatcgtgcgaagaagctgggtgaagaggtgcgccgtatgattaacaataaggacaccaagttgctgaccacgctggaactgatcg acgatattgagcgtctgggtctgggttaccgcttcaaagaagagatcatgcgcgcgttggaccgtttcgttacgctgaagggttgtgaagagt tcacgaacggttccatccatgatacggcgttgagcttccgtttgctgcgtcagcacggttttggcgtgagccaggacatgtttaactgtttcaa agaccagaaaggcaatttcaaagagtgtctgtctaaagacatcaagggtctgctgagcctgtatgaggcgagctatctgggttttgagggtg agaatctgctggatgaagcacgtgagttcactaccatgcacctgaaagacctgaaaggcgacgtttcgcgtaccctgaaagaagaggtcc gtcacagcctggagatgccgctgcatcgtcgcatgcgtcgtctggagcagcgttggtatatcgatgcttacaacatgaaagaagcccatgat cgtaagctgctggagctggcaaaattggacttcaacattgttcagagcgtgcaccagcgcgacttgaaggacatgagccgttggtggcaag aaatgggcctgggcaacaagctgtcgttcgcacgcgaccgcctgatggagtgcttcttttttctccgtcggtatggcgtttgagccgcaattta gcaatagccgcaaagcggtcaccaagatgtttagctttatcaccgtgatcgatgacatctacgatgtttacgcgaccctggaagagctggaa atgttcacggatatcgtgcaacgctgggacgtgaaagcagtcaaagacttgccggagtatatgaagttgtgctttctggccttgtttaacaccg

tcaatgagatggtttacgacacgctgaaagagcaaggcgtcgacattctgccgtacctgaccaaggcctggggtgatatttgcaaagccttc
ctgcaagaaaccaagtggcgctattacaagcgtaccccgagcagcgaggattacctggacaacgcatggatttccgttagcggtgctctgc
tgctgattcacgcgtacttcctgatgtctccgagcatcaccgatcgtgcgctgaaaggcctggaagattatcataacatcctgcgttggccga
gcattatctttcgtctgaccaacgatttgggtactagcacggctgagctggagcgtggcgaaaccgccaatagcattttgtgttatatgcgcga
aaccagccgctctgaggattttgcgcgtgaacacatcagcaatctgattgataaaacctggaagaagatgaataaagatcgtttcagcgaca
gcccattcgaagaaccgtttctggaaacggcaattaacctggcccgcatcagccattgcatctaccagcacggcgatggtcacggtgcgcc
agacacgcgcaccaaggaccgtgttctgtctctgatcattgagccgattccgtgctacgatcctagcacgaatttccatagccagattcacct
gtaa (SEQ ID NO:14)

*Amino Acid Sequence of P. montana IspS*

MNSRRSANYQPNLWNFEFLQSLENDLKVEKLEEKATKLEEEVRCMINRVDTQPLSLLEL
IDDVQRLGLTYKFEKDIIKALENIVLLDENKKNKSDLHATALSFRLLRQHGFEVSQDVFE
RFKDKEGGFSGELKGDVQGLLSLYEASYLGFEGENLLEEARTFSITHLKNNLKEGINTK
VAEQVSHALELPYHQRLHRLEARWFLDKYEPKEPHHQLLLELAKLDFNMVQTLHQKE
LQDLSRWWTEMGLASKLDFVRDRLMEVYFWALGMAPDPQFGECRKAVTKMFGLVTII
DDVYDVYGTLDELQLFTDAVERWDVNAINTLPDYMKLCFLALYNTVNDTSYSILKEKG
HNNLSYLTKSWRELCKAFLQEAKWSNNKIIPAFSKYLENASVSSSGVALLAPSYFSVCQ
QQEDISDHALRSLTDFHGLVRSSCVIFRLCNDLATSAAELERGETTNSIISYMHENDGTSE
EQAREELRKLIDAEWKKMNRERVSDSTLLPKAFMEIAVNMARVSHCTYQYGDGLGRP
DYATENRIKLLLIDPFPINQLMYV (SEQ ID NO:15)

*Nucleic Acid Sequence of P. montana IspS*

atgaattcccgtcgttccgcaaactatcagccaaacctgtggaatttcgaattcctgcaatccctggagaacgacctgaaagtggaaaagct
ggaggagaaagcgaccaaactggaggaagaagttcgctgcatgatcaaccgtgtagacacccagccgctgtccctgctggagctgatcg
acgatgtgcagcgcctgggtctgacctacaaatttgaaaaagacatcattaaagccctggaaaacatcgtactgctggacgaaaacaaaaa
gaacaaatctgacctgcacgcaaccgctctgtctttccgtctgctgcgtcagcacggtttcgaggtttctcaggatgtttttgagcgtttcaagg
ataaagaaggtggtttcagcggtgaactgaaaggtgacgtccaaggcctgctgagcctgtatgaagcgtcttacctgggtttcgagggtga
gaacctgctggaggaggcgcgtacctttccatcacccacctgaagaacaacctgaaagaaggcattaataccaaggttgcagaacaagt
gagccacgccctggaactgccatatcaccagcgtctgcaccgtctggaggcacgttggttcctggataaatacgaaccgaaagaaccgca
tcaccagctgctgctggagctggcgaagctggattttaacatggtacagaccctgcaccagaaagagctgcaagatctgtcccgctggtgg
accgagatgggcctggctagcaaactggattttgtacgcgaccgcctgatggaagtttatttctgggcactgggtatggcgccagacccgc
agtttggtgaatgtcgcaaagctgttactaaaatgtttggtctggtgacgatcatcgatgacgtgtatgacgtttatggcactctggacgaactg
caactgttcaccgatgctgtagagcgctgggacgttaacgctattaacaccctgccggactatatgaaactgtgtttcctggcactgtacaac
accgttaacgacacgtcctattctattctgaaagagaaaggtcataacaacctgtcctatctgacgaaaagctggcgtgaactgtgcaaagcc
tttctgcaagaggcgaaatggtccaacaacaaaattatcccggctttctccaagtacctggaaaacgccagcgtttcctcctccggtgtagcg
ctgctggcgccgtcttacttttccgtatgccagcagcaggaagacatctccgaccacgcgctgcgttccctgaccgacttccatggtctggtg
cgttctagctgcgttatcttccgcctgtgcaacgatctggccacctctgcggcggagctggaacgtggcgagactaccaattctatcattagc
tacatgcacgaaaacgatggtaccagcgaggaacaggcccgcgaagaactgcgtaaactgatcgacgccgaatggaaaaagatgaatc
gtgaacgcgttagcgactccaccctgctgcctaaagcgttcatggaaatcgcagttaacatggcacgtgtttcccactgcacctaccagtatg
gcgatggtctgggtcgcccagactacgcgactgaaaaccgcatcaaactgctgctgattgacccttcccgattaaccagctgatgtatgtct
aa (SEQ ID NO:16)

*Nucleic acid sequence of pCL201*

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcc
ctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttc
cgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcc
ctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataaggg
attttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtgg
cacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcat
cgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccg
aggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaa
taaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacagg
ccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgct

gttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatatt
cttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaa
gaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggc
gcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgtt
ggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccccttgtattactgtttatgtaagcagacagttttattgtt
catgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcg
cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaact
ggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatac
ctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataag
gcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtg
agctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacg
agggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcag
gggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgtta
tcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtga
gcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaat
ctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccg
ctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttca
ccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgt
ccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcct
ccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcc
cggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttc
gttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcg
tttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctc
gcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccg
tggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgca
agattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccgg
cacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactg
ggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttcc
agtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttc
ttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccag
caggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcacc
aacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccct

cattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatat
ttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccag
atgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcg
ccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgg
gaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggc
gctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcag
cccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacg
gggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatata
ggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaat
acgactcactatagggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatatacatatggaagcac
gtcgctctgcgaactacgaacctaacagctgggactatgattacctgctgtcctccgacacggacgagtccatcgaagtatacaaagacaa
agcgaaaaagctggaagccgaagttcgtcgcgagattaataacgaaaaagcagaatttctgaccctgctggaactgattgacaacgtccag
cgcctgggcctgggttaccgtttcgagtctgatatccgtggtgcgctggatcgcttcgtttcctccggcggcttcgatgcggtaaccaagactt
ccctgcacggtacggcactgtctttccgtctgctgcgtcaacacggttttgaggtttctcaggaagcgttcagcggcttcaaagaccaaaacg
gcaacttcctggagaacctgaaggaagatatcaaagctatcctgagcctgtacgaggccagcttcctggctctggaaggcgaaaacatcct
ggacgaggcgaaggttttcgcaatctctcatctgaaagaactgtctgaagaaaagatcggtaaagagctggcagaacaggtgaaccatgc
actggaactgccactgcatcgccgtactcagcgtctggaagcagtatggtctatcgaggcctaccgtaaaaaggaggacgcgaatcaggtt
ctgctggagctggcaattctggattacaacatgatccagtctgtataccagcgtgatctgcgtgaaacgtcccgttggtggcgtcgtgtgggt
ctggcgaccaaactgcactttgctcgtgaccgcctgattgagagcttctactgggccgtgggtgtagcattcgaaccgcaatactccgactg
ccgtaactccgtcgcaaaaatgttttctttcgtaaccattatcgacgatatctacgatgtatacggcaccctggacgaactggagctgtttactg
atgcagttgagcgttgggacgtaaacgccatcaacgacctgccggattacatgaaactgtgctttctggctctgtataacactattaacgaaat
cgcctacgacaacctgaaagataaaggtgagaacatcctgccgtatctgaccaaagcctgggctgacctgtgcaacgctttcctgcaagaa
gccaagtggctgtacaacaaatctactccgacctttgacgactacttcggcaacgcatggaaatcctcttctggcccgctgcaactggtgttc
gcttacttcgctgtcgtgcagaacattaaaaaggaagagatcgaaaacctgcaaaaataccatgacaccatctctcgtccttcccatatcttcc
gtctgtgcaatgacctggctagcgcgtctgcggaaattgcgcgtggtgaaaccgcaaatagcgtttcttgttacatgcgcactaaaggtatct
ccgaagaactggctaccgaaagcgtgatgaatctgatcgatgaaacctggaaaaagatgaacaaggaaaaactgggtggtagcctgttcg
cgaaaccgttcgtggaaaccgcgatcaacctggcacgtcaatctcactgcacttatcataacggcgacgcgcatacctctccggatgagct
gacccgcaaacgcgttctgtctgtaatcactgaaccgattctgccgtttgaacgctaaggatccgaattcgagctccgtcgacaagcttgcg
gccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctg

agcaataactagcataacccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat (SEQ ID NO:34)

*Nucleic acid sequence of pEWL1036*

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaat
cactgcataattcgtgtcgctcaaggcgcactcccgttctggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatga
gctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagcgccgctgagaaaaagc
gaagcggcactgctctttaacaatttatcagacaatctgtgtgggcactcgaccggaattatcgattaactttattattaaaaattaaagaggtat
atattaatgtatcgattaaataaggaggaataaaccatgaacacccgtcgcagcgccaattaccagccgaacctgtgggatttcgagtttttgc
aaagcgtcgaaaacgacttgcaggtggaacgcctggaagagcgtgcacgtaagctggaagaagaagtgcgtggtctgatgaaaaaggt
cgagattgagccgttgagcctgctggaactgatggacaacgttgagcgcctgggtctgacctacaaattcgaagaggacatcaaaagcgc
gttgaataaccgcattgtcccactgttgcatcaccatactatcaataagtacggtctgcacgccacggctctgagcttccgtttcctgcgtcaac
acgcctttcacgtcagcccggatgtttttgaaagcttcaaagaagaaggtaagttcaagaaagagattagcggcgacgtgctgggtctgctg
aacctgtacgagactagctacctgggctttgaaggtgaaaccattctggacgaggcacgcgccttcagcgctacccatctgaaaaatctgtt
gcaaaccaaccaggtgcagaataaagttatggcggagaaggtccgccatgcgttggagctgccgtatcaccgtcgtgttcaccgtttggaa
gcccgctggtttattgagcgctatgagcagaaagaggcgcatgatggtgccttgctggagctggcgaaactggatttcaacatggttcagtc
tgtgatgaagaaagagctgcaagagctgagccgctggtggcgtgagatcggtctgaccagcaagctggacttcgtgcgtgatcgtctgat
ggaagtgtacttttgggcgctgggtatggctccgcacccgcagctgacggagtgccgtaaagcagtgaccaagatgtttggcctggttacc
atcattgacgatgtttacgatgtgtatggcaccctggacgagctgcaactgtttacggatgcggttgaccgttgggacgttaacgcagtcgaa
acgctgccggactacatgaaactgtgttacctggcgctgtataactccgttaatgacacggcatatagcactctgcgtgagaagggtgacaa
tagcctgcctcacttggcaaagtcgtggcgtgatctgtgtaaggcgtttctgcaagaggcgaagtggagcaataacaagatcattccgccgt
tcgatgcgtacatccgcaacgcatctgtcagcagcggcggtgctctgttggcgccatgttacttctccgttacgcaagacagcacgag
ccaggccatcgattctattacgaactaccacggcatcgtccgttcgagctgcgcaatcttccgcctgtgcaatgacctggcgacctctgctgc
ggagctggagcgtggcgaaaccaccaattccatcacgtcctatatgaccgaaaatggcaccaccgaagaagaggcgcgtgaaagcctg
ggtaaactgattgaccaagagtggaagaaaatgaatcgtgatgtggtcctggaaagcgcgtatccgaacgtgtttaaagaaattgcgattaa
catggcacgcgttagccattgcacctatcagtatggcgatggtctgggtcgtccggatgatactgcggagaatcgtatcaagctgtctctgat
cgaaccgattccgatcaactaactgcataaaggaggtaaaaaaacatggtatcctgttctgcgccgggtaagatttacctgttcggtgaacac
gccgtagtttatggcgaaactgcaattgcgtgtgcggtggaactgcgtacccgtgttcgcgcggaactcaatgactctatcactattcagagc
cagatcggccgcaccggtctggatttcgaaaagcacccttatgtgtctgcggtaattgagaaaatgcgcaaatctattcctattaacggtgtttt
cttgaccgtcgattccgacatcccggtgggctccggtctgggtagcagcgcagccgttactatcgcgtctattggtgcgctgaacgagctgt
tcggctttggcctcagcctgcaagaaatcgctaaactgggccacgaaatcgaaattaaagtacagggtgccgcgtccccaaccgatacgta
tgtttctaccttcggcggcgtggttaccatcccggaacgtcgcaaactgaaaactccggactgcggcattgtgattggcgataccggcgtttt

ctcctccaccaaagagttagtagctaacgtacgtcagctgcgcgaaagctacccggatttgatcgaaccgctgatgacctctattggcaaaa
tctctcgtatcggcgaacaactggttctgtctggcgactacgcatccatcggccgcctgatgaacgtcaaccagggtctcctggacgccctg
ggcgttaacatcttagaactgagccagctgatctattccgctcgtgcggcaggtgcgtttggcgctaaaatcacgggcgctggcggcggtg
gctgtatggttgcgctgaccgctccggaaaaatgcaaccaagtggcagaagcggtagcaggcgctggcggtaaagtgactatcactaaac
cgaccgagcaaggtctgaaagtagattaaagtctagttaaagtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagc
ctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatg
ccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaac
gaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcggattt
gaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcct
gacggatggccttttgcgtttctacaaactcttttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgc
ttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcaccc
agaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcctt
gagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaag
agcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagta
agagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccg
ctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacacc
acgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatg
gaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtc
tcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa
cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaa
cttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcaga
ccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtg
gtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagcc
gtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataa
gtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagccca
gcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgga
caggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcg
ggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttac
ggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata
ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcat
ctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgt
gactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagaca

agctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaa
ggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaacctttcgcggtatggcatgatagcgcccggaagagagtcaattca
gggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggcc
agccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggc
gggcaaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgcc
gatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgca
acgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttctt
gatgtctctgaccagacacccatcaacagtattattttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcac
cagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattc
agccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactg
cgatgctggttgccaacgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtag
tgggatacgacgataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgt
ggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggc
gcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtg
agcgcaacgcaattaatgtgagttagcgcgaattgatctg (SEQ ID NO:32)

## CLAIMS

1. An isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to SEQ ID NO: 1, wherein said polypeptide has one or more amino acid residue(s) corresponding to one or more amino acid residue(s) corresponding to SEQ ID NO:1, wherein said one or more amino acid residue(s) are selected from the group consisting of F287, G397, N438, E451, and Y514, and wherein said isolated polypeptide is not from *P. montana*.

2. The isolated polypeptide of claim 1, wherein the isolated polypeptide is an isoprene synthase selected from the group consisting of *Arachis sp.*, *Mucana sp.*, *Cajanus sp.*, *Glycine sp.*, *Lotus sp.*, and *Medicago sp.*

3. The isolated polypeptide of claim 2, wherein the isolated polypeptide is an isoprene synthase selected from the group consisting of *A. hypogaea*, *M. pruriens, C. cajans, G. max, G. soja, L. japonicus,* and *M. truncatula.*

4. The isolated polypeptide of claim 2, wherein the isolated polypeptide is an *A. hypogaea* isoprene synthase.

5. The isolated polypeptide of claim 2, wherein the isolated polypeptide is an *M. pruriens* isoprene synthase.

6. The isolated polypeptide of any one of claims 1-5, wherein said isolated polypeptide has a reduced substrate inhibition as compared to a poplar isoprene synthase.

7. The isolated polypeptide of any one of claims 1-6, wherein said isolated polypeptide has increased isoprene synthase activity as compared to a poplar isoprene synthase.

8. The isolated polypeptide of claim 7, wherein increased isoprene synthase activity is indicated by a host cell comprising the isoprene synthase displaying improved growth in the presence of mevalonic acid compared to a host cell comprising a poplar isoprene synthase.

9. The isolated polypeptide of any one of claims 1-8, wherein the isolated polypeptide is an isoprene synthase selected from the group consisting of: *A. hypogaea* (SEQ ID

NO:3), *G. max* 1 (SEQ ID NO:5), *G. max* 2 (SEQ ID NO:7), *M. pruriens* (SEQ ID NO:9), and *C. cajans* (SEQ ID NO:11).

10. The isolated polypeptide of claim 9, wherein the isolated polypeptide is an *A. hypogaea* isoprene synthase comprising the amino acid sequence of SEQ ID NO:3.

11. An isolated polypeptide from a legume having isoprene synthase activity comprising at least 70% sequence identity to a legume isoprene synthase comprising the amino acid sequence of SEQ ID NO: 3.

12. An isolated polypeptide from a legume having isoprene synthase activity of any one of claims 1-11, wherein said isoprene synthase has a $K_{cat}$ value of at least about 1.3.

13. An isolated polypeptide from a legume having isoprene synthase activity of any one of claims 1-11, wherein said isoprene synthase has a $K_m$ value of at least about 2.5.

14. An isolated polypeptide from a legume having isoprene synthase activity of any one of claims 1-11, wherein said isoprene synthase has a $K_i$ value of at least about 13.0.

15. An isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to a poplar isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1, and wherein said polypeptide has a $K_{cat}$ value of at least about 1.3.

16. An isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to a poplar isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1, and wherein said polypeptide has a $K_m$ value of at least about 2.5.

17. An isolated polypeptide from a legume having isoprene synthase activity, wherein said polypeptide comprises at least 40% sequence identity to a poplar isoprene synthase comprising the amino acid sequence of SEQ ID NO: 1, and wherein said polypeptide has a $K_i$ value of at least about 13.0.

18. A host cell comprising a heterologous polynucleotide sequence encoding an isoprene synthase of any one of claims 1-17 in operable combination with a promoter.

19. The host cell of claim 18, wherein the polynucleotide sequence is contained within a plasmid.

20. The host cell of claim 18, wherein the polynucleotide sequence is integrated into a chromosome of the host cell.

21. The host cell of claim 18, wherein the host is selected from the group consisting of gram-positive bacterial cells, gram-negative bacterial cells, filamentous fungal cells, and yeast cells.

22. The host cell of claim 18, wherein the host is selected from the group consisting of Escherichia sp. (*E. coli*), Panteoa sp. (*P. citrea*), Bacillus sp. (*B. subtilis*), Yarrowia sp. (*Y. lipolytica*), Trichoderma (*T. reesei*) and *Saccharomyces* (*S. cerevisiae*).

23. The host cell of claim 18, wherein the host cell is cultured in a medium comprising a carbon source selected from the group consisting of glucose, glycerol, glycerine, dihydroxyacetone, yeast extract, biomass, molasses, sucrose, and oil.

24. The host cell of claim 18, wherein the host cell further comprises one or more heterologous or native nucleic acid(s) encoding one or more mevalonate (MVA) pathway polypeptides.

25. The host cell of claim 24, wherein the host cell further comprises a heterologous nucleic acid encoding a mevalonate (MVA) pathway polypeptide selected from the group consisting of an MVA pathway polypeptide from *Saccharomyces cerevisiae* and *Enterococcus faecalis*.

26. The host of claim 24, further comprising one or more nucleic acids endoding a heterologous or native nucleic acid encoding an IDI polypeptide.

27. The host cell of claim 18, wherein the host cell further comprises a heterologous or native nucleic acid encoding an IDI polypeptide and/or a heterologous or native nucleic acid encoding a DXS polypeptide, optionally in combination with the native DXP pathway.

28. The host cell of claim 27, wherein the host cell further comprises one or more nucleic acids encoding an IDI polypeptide and the DXS polypeptide.

29. The host cell of claim 27, wherein the host cell comprises one vector encoding the isoprene synthase, the IDI polypeptide, and the DXS polypeptide.

30. The host cell of claim 18, wherein the host cell further comprises one or more nucleic acids encoding an MVA pathway polypeptide and/or a DXS polypeptide.

31. The host cell of claim 30, wherein the host cell further comprises one or more nucleic acids encoding a DXS polypeptide, an IDI polypeptide, or one or more of the rest of the DXP pathway polypeptides, and a MVA pathway polypeptide.

32. The host cell of claim 26 further comprising DXS.

33. A method of producing isoprene, comprising:

(a) culturing the host cells of any one of claims 18-32 under suitable culture conditions for production of isoprene; and

(b) producing the isoprene.

34. The method of claim 33 further comprising (c) recovering the isoprene.

35. The method of claim 34 further comprising (d) polymerizing isoprene.

36. A method of detecting isoprene synthase activity, comprising: (a) culturing a host cell comprising the expression vector under conditions suitable for producing a legume isoprene synthase; (b) lysing the host cells with a lysis buffer comprising lysozyme to produce a cell lysate; and (c) detecting isoprene synthase activity in the cell lysate by measuring isoprene production from dimethylallyl diphosphate (DMAPP).

37. The method of claim 36, wherein the host cell is selected from the group consisting of gram-positive bacterial cells, gram-negative bacterial cells, filamentous fungal cells, and yeast cells.

38. The method of claim 37, wherein the host cell is selected from the group consisting of *Escherichia sp. (E. coli), Panteoa sp. (P. citrea), Bacillus sp. (B. subtilis), Yarrowia sp. (Y. lipolytica), Trichoderma (T. reesei)*, and *Saccharomyces (S. cerevisiae)*.

39. The method of claim 36, wherein the host cell is cultured in a medium that includes a carbon source selected from the group consisting of glucose, glycerol, glycerine, dihydroxyacetone, yeast extract, biomass, molasses, sucrose, and oil.

40. A method of detecting isoprene in a plurality of samples in a high-throughput screening comprising:

(a) providing: i) a plurality of samples each comprising isoprene synthase; ii) a glass plate comprising a plurality of wells; and iii) a seal for the glass plate;

(b) placing the plurality of samples in the plurality of wells of the glass plate;

(c) sealing the glass plate with the seal to produce a sealed glass plate having a headspace associated with the sample in each of the plurality of wells;

(d) incubating the glass plate under conditions in which the isoprene synthase is active; and

(e) detecting isoprene in the headspace.

41. The method of claim 40, wherein the isoprene is detected by gas chromatography-mass spectrometry (GC-MS).

42. The method of claim 40, wherein the plurality of samples comprise host cells comprising an expression vector comprising a polynucleotide sequence encoding an isoprene synthase variant in operable combination with a promoter.

43. The method of claim 40, wherein the plurality of samples comprise a lysate of the host cells, lysozyme, and dimethylallyl diphosphate (DMAPP).

44. The method of claim 40, wherein the glass plate is a deep-well glass block.

45. The method of claim 40, wherein the plurality of wells comprises at least 24 wells.

46. The method of claim 40, wherein the plurality of wells each comprise a volume of 2 ml or less.

# Figure 1

peanut_fom_1n24. TRRSANYQPNLWDFEFLQSVENDLQVERLEERARKLEEEVRGLMKKVEIEPLSLLELMDN
Palba_from_1n24. ARRSANYEPNSWDYDYLLSSDTDESIEVYKDKAKKLEAEVRREINNEKAEFLTLLELIDN

peanut_fom_1n24. VERLGLTYKFEEDIKSALNNRIVPLLHHHTINKYG----LHATALSFRFLRQHAFHVSPD
Palba_from_1n24. VQRLGLGYRFESDIRGALDR-----FVSSGGFDAVTKTSLHGTALSFRLLRQHGFEVSQE

peanut_fom_1n24. VFESFKEEGK-FKKEISGDVLGLLNLYETSYLGFEGETILDEARAFSATHLKNLLQTNQV
Palba_from_1n24. AFSGFKDQNGNFLENLKEDIKAILSLYEASFLALEGENILDEAKVFAISHLKELSE-EKI

peanut_fom_1n24. QNKVMAEKVRHALELPYHRRVHRLEARWFIERYEQKEAHDGALLELAKLDFNMVQSVMKK
Palba_from_1n24. GKELAEQ-VNHALELPLHRRTQRLEAVWSIEAYRKKEDANQVLLELAILDYNMIQSVYQR

peanut_fom_1n24. ELQELSRWWREIGLTSKLDFVRDRLMEVYFWALGMAPHPQLTECRKAVTKMFGLVTIIDD
Palba_from_1n24. DLRETSRWWRRVGLATKLHFARDRLIESFYWAVGVAFEPQYSDCRNSVAKMFSFVTIIDD

peanut_fom_1n24. VYDVYGTLDELQLFTDAVDRWDVNAVETLPDYMKLCYLALYNSVNDTAYSTLREKGDNSL
Palba_from_1n24. IYDVYGTLDELELFTDAVERWDVNAINDLPDYMKLCFLALYNTINEIAYDNLKDKGENIL

peanut_fom_1n24. PHLAKSWRDLCKAFLQEAKWSNNKIIPPFDAYIRNASVSSSGGALLAPCYFSVTQDSTSQ
Palba_from_1n24. PYLTKAWADLCNAFLQEAKWLYNKSTPTFDDYFGNAWKSSSGPLQLVFAYFAVV-QNIKK

peanut_fom_1n24. -AIDSITNYHGIVRSSCAIFRLCNDLATSAAELERGETTNSITSYMTENGTTEEEARESL
Palba_from_1n24. EEIENLQKYHDTISRPSHIFRLCNDLASASAEIARGETANSVSCYMRTKGISEELATESV

peanut_fom_1n24. GKLIDQEWKKMNRDVVLESAYPNVFKEIAINMARVSHCTYQYGDGLGRPDDTAENRIKLS
Palba_from_1n24. MNLIDETWKKMNKEKLGGSLFAKPFVETAINLARQSHCTYHNGDAHTSPDELTRKRVLSV

peanut_fom_1n24. LIEPIP
Palba_from_1n24. ITEPIL

Figure 2A


1 50
A_hypogaea (1) METPRSANYQPNLWDFEFLQ-SVEND-----LQVERLEEPARKLEEEVRG
Glyma09g21900.1 (1) ----------PARKVEEEVRR
Glyma20g18280_1_FG (1) METRRSANYQPNLWNFEFLPPSLEND-----HKVEKLEEPAKKVEEEVRK
P_montana (1) METPRSANYQPNLWNFEFLQ-SLEND-----LKVEKLEEKAEKLEEEVRC
C_limon (1) METPRSANYQPSIWDHDFLQSLNSN------YTDETERPKAEELKGKVKI
Q_ilex (1) METRRSANYQPSIWNHDYIPSLRIE------YVGETCTRQINVEKEQVRM
V_vinifera (1) METPRTANYHPTIWDYDYVQSLRSD------YVGETETRELDKLKPDVKP
C_tenuipile (1) METPRSGNYEPSIWDYDFVQSLGSG------YKVEAHGTPVKKLKEVVKH
E.globulus_trunc (1) MEGRRSANYQPSVWTYNYLQSIVAGEGRQSRREVEQQKEKVQILEEEVRG
M_alternifolia (1) METPRSANYQPSVWTYNYLQSLVADDIRRSRPEVEQEPEKAQILEEDVRG
Glyma06g45780.1 (1) ----------DIRYEDKAKKLLEEVRR
Glyma12g10990.1 (1) ----KSANYQPNLWNYDFLQSLKND------YADVKYEIMAKK-LEEVRR
L_japonicus (1) METPKSANYQPNIWNYDYLQSLKLG------YADAHYEDMAKKLQEEVRR
M.truncatula (1) ----------EEVKR
Q_petraea (1) METRQSANFQPSLWSYEYLQSLKNG------YEADLIEDPAKKLGEEVRR
P.alba (1) MEAPRSANYEPNSWDYDYLLSSDTD------ESIEVYKDKAKKLEAEVRR
S.alba (1) METPRSANYEPNSWDYDYLLSSDTE------DAIEVYKDKAKKLDAEVRS
Consensus (1) METRRSANYQP IW YDYLQSL E Y DPAKKL EEVR
51 100
A_hypogaea (45) LMKKVEIEPLSLLELMDNVERLGLTYKFEEDIKSALNNRIVPLLHHHTIN
Glyma09g21900.1 (12) MINGADTEALRLLELIDEIQRLGLTYKFEKDIFKALEK--TISLDENEKH
Glyma20g18280_1_FG (46) VINGIDTKPD-LLELIDDVQHLGLTYKFEKDIIKALEK--IVSLDENEEH
P_montana (45) MINRVDTQPLSLLELIDDVQRLGLTYKFEKDIIKALEN--IVLLDENKKN
C_limon (45) AIKDVT-EPLDQLELIDNLQRLGLAYRFETEIRNILHNIYNN-NKDYVWR
Q_ilex (45) MLHKVV-NPLEQLELIEILQRLGLSYHFEEEIKRILDGVYNNDHGGDTWK
V_vinifera (45) MLGKVK-KPLDQLELIDVLQRLGIYYHFKDEIKPILNSIYNQYNRHEEWQ
C_tenuipile (45) LLKETDS-SLAQIELIDKLFRLGLPWLFKNEIKQVLYTISSDNTSIEMRK
E.globulus_trunc (51) ALNDEKAETFTIFATVDDIQRLGLGDHFEEDISNALRPCVSKGAVFMSLQ
M_alternifolia (51) ALNDGNAEPMAIFALVDDIQRLGLGRYFEEDISKALRPCLSQYAVTGSLQ
Glyma06g45780.1 (18) MIKDENTDIWIKLELIDDVERLGIGYSFDMEIGEALHRCLSSETFIDTIT
Glyma12g10990.1 (40) MIKDENSEIWVTLDLIDNVKRLGLSYHFDKEIREALHRFLSLEFCN--AT
L_japonicus (45) LIKDDKAEIWTTLELIDDVKRLGLGYHFEKEIREVLNKFLSLNTCV----
M.truncatula (6) MIKDEN--VN-ILELIDTVQQLGLSYHFEEEIGEALDRFLSLEKCSG-RN
Q_petraea (45) MINNKDTKLITTLELIDDIERLGLGYRFAEEIMRALDFFVTLKGCEEFTN
P.alba (45) EINNEKAEFLTLLELIDNVQRLGLGYRFESDIRGALDRFVSSGGFDAVTK
S.alba (45) KINNETAEFLTQLELIDTIQRLGLGYRFESDIRRALDRFVSSGGFEAVAK
Consensus (51) MI E L LELID VQRLGL Y FE EIR AL F LS

Figure 2B

Figure 2C


201 250
A_hypogaea (191) ELPYHRRVHRLEAPWFIERYEQKEAHDGALLELAKLDFNMVQSVMKKELQ
Glyma09g21900.1 (155) ELPYHRGLNRLEAPWFLEKYEPNESHHHVLLELAKLDFNLVQVMYQKELK
Glyma20g18280_1_FG (188) ELPYHRPLQRLEAPWYLEKYEPKEPHHQLLLELAKLDFNMVQLLHQKELQ
P_montana (188) ELPYHQRLHRLEAPWFLDKYEPKEPHHQLLLELANLDFNMVQTLHQKELQ
C_limon (188) ELPLHWKVPMLEAPWFIHVYEKREDKNHLLLELARMEFNTLQAIYQEELK
Q_ilex (189) EFPLHWRMPRLEAPWFINIYRHNQDVNPLLLEFAELDFNIVQAAHQADLK
V_vinifera (189) ELPLHWRMPRLEARWFIDVYEKRQDMNPLLLEFAKLDFNMVQATHQEDLR
C_tenuipile (186) DMPLHWPLEKLEAPWYMDIYMREEGMNSSLLELAMLHFNIVQTTFQTNLK
E.globulus_trunc (194) ELPLHRPMPLLEARRSIEAYSRREYTNPQTLELALTDFNVSQSTLQRDLQ
M_alternifolia (194) ELPLHRRMPLLEARRFIEAYSRRSNVNPRILELAVMKFNSSQLTLQRDLQ
Glyma06g45780.1 (166) ELPLHHRIQRLEAPWYIESYAKRKDANMVLLEAAKLDFNIVQSTLQTDLQ
Glyma12g10990.1 (186) ELPLHHRIQRLEAPWYIESYAKRKDANWVLLEAAKLDFNIVQSTLQKDLQ
L_japonicus (186) ELPLHRPIHRLEAPWYTLSYSRRKDANWVLLEAAKLDFNMVQSTLQKDLQ
M.truncatula (148) ELPLYRPVQSLEARWFIDSYENRKDANKVLLEAAKLNFNIVQSTLQQDLK
Q_petraea (188) EMPLHRRMRRLEQRWYIDAYNMKEAHDRKLLELAKLDFNIVQSVHQRDLK
P.alba (189) ELPLHRRTQRLEAVWSIEAYPKKEDANQVLLELAILDYNMIQSVYQRDLR
S.alba (189) ELPLHRRTQRLEAVWSIEAYPKKEDANRVLLELAILDYNMVQSVYQRDLR
Consensus (201) ELPLHRRM RLEARWFIEAY KRED N VLLELAKLDFNMVQS HQKDLK
251 300
A_hypogaea (241) ELSRWWREIGLTSKLDFVRDRLMEVYFWALGMA-PHPQLTECRKAVTKMF
Glyma09g21900.1 (205) ELSRWWSEMGLTSKLKFVRDRLMEVYFWVLGMA-PRPQFSECRKAVTKTF
Glyma20g18280_1_FG (238) ELSRWWSEMGLASKLEFARDRLMEVYFWALGMA-PDPQFRECRKAVTKMF
P_montana (238) DLSRWWTEMGLASKLDFVRDRLMEVYFWALGMA-PDPQFGECRKAVTKMF
C_limon (238) EISGWWKDTGLGEKLSFARNRLVASFLWSMGIA-FEPQFAYCRRVLTISI
Q_ilex (239) QVSTWWKSTGLVENLSFARDRPVENFFWIVGLI-EQPQFGYCRRMFTKVF
V_vinifera (239) HMSSWWSSTRIGEKLNFARDRLMENFLWTVGVI-FEPQYGYCRRMSTKVN
C_tenuipile (236) SLSRWWKDLGLGEQLSFTRDRLVECFFWAAAMT-PEPQFGFCQEVVAKVA
E.globulus_trunc (244) EMLGWWNNTGLAKRLSFARDRLIECFFWAVGIA-HEPSLSICRKAVTKAF
M_alternifolia (244) DMLGWWNNVGLAKRLSFARDRLMECFFWAVGIA-REPALSNCRKGVTKAF
Glyma06g45780.1 (216) EMSRWWKGMGLASKLSFSRDRLMECFFWIVGMV-FEPQLSDLRKGLTKVA
Glyma12g10990.1 (236) EMSPWWKRMGLAPKLSFSRDRLMECFFWSMGMA-FEPQFSDLPKGLTKVT
L_japonicus (236) EMSPWWKGMGLAPKLSFSRDRLMECFFWIVGMA-FEPKYSDLPKGLTKVT
M.truncatula (198) EMSPWWKGMGLAPRLSFGRDRLMECFFWAAGMTPFEPQFSNIPKGLTKVC
Q_petraea (238) DMSRWWQEMGLGNKLSFARDRLMECFFFSVGMA-FEPQFSNSPKAVTKMF
P.alba (239) ETSRWWRPVGLATKLHFARDRLIESFYWAVGVA-FEPQYSDCRNSVAKMF
S.alba (239) ETSRWWRPVGLATKLHFARDRLIESFYWAVGVA-FEPQYSDCRLSVAKMF
Consensus (251) EMSRWWK MGLA KLSFARDRLME FFWAVGMA FEPQFSDCRKAVTKVF

Figure 2D


301 350
A_hypogaea (290) GLVTIIDDVYDVYGTLDELQLFTDAVDRWDVN-AVETLPDYMKLCYLALY
Glyma09g21900.1 (254) ALIGIIDDVYDVYGTLDELQLFTDAIERWDVN-AMNTLPDYMKLCYLAVY
Glyma20g18280_1_FG (287) GLVTIIDDVYDIYGTLDELQLFTDAVERWDVN-VVNTLPDYMKLCYLALY
P_montana (287) GLVTIIDDVYDVYGTLDELQLFTDAVERWDVN-AINTLPDYMKLCFLALY
C_limon (287) ALITVIDDIYDVYGTLDELELFTDAVARWDINYALKHLPGYMKMCFLALY
Q_ilex (288) ALITTIDDVYDVYGTLDELELFTDVVERWDIN-AMDQLPDYMKICFLTLH
V_vinifera (288) TLITTIDDVYDVYGTMDELELFTDVVDRWDIN-AMDFLPEYMKICFLALY
C_tenuipile (285) QLIIIIDDIYDVYGTVDELELFTNAIDRWDLE-AMEQLPEYMKTCFLALY
E.globulus_trunc (293) ALILVLDDVYDVFGTLEELELFTEAVRRWDLN-AVEDLPVYMKICYLALY
M_alternifolia (293) SLILVLDDVYDVFGTLDELELFTDAVPRWHED-AVENLPGYMKLCFLALY
Glyma06g45780.1 (265) SLITTIDDVYDVYGTLDELELFTAAVESWDVK-AVQVLPDYMKICFLALY
Glyma12g10990.1 (285) SLITTIDDVYDVYGSLDELELFTKAVESWDIK-AVQVMPEYMKICFLALY
L_japonicus (285) SLITTIDDIYDVHGTLEELELFTAIVESWDIK-AMQVLPEYMKISFLALY
M.truncatula (248) SLITLIDDIYDVYGTLDELELFTTAVESWDIN-AIQILPEYMKIFFLALY
Q_petraea (287) SFITVIDDIYDVYATLEELEMFTDIVQRWDVK-AVKDLPEYMKLCFLALF
P.alba (288) SFVTIIDDIYDVYGTLDELELFTDAVERWDVN-AINDLPDYMKLCFLALY
S.alba (288) SFVTIIDDIYDVYGTLEELELFTDAVERWDVS-AIDDLPDYMKLCFLALY
Consensus (301) SLITIIDDVYDVYGTLDELELFTDAVERWDVN AV LPDYMKLCFLALY
351 400
A_hypogaea (339) NSVNDTAYSTLREKGDNSLPHLAKSWRDLCKAFLQEAKWSNNKITPPFDA
Glyma09g21900.1 (303) NTVNDTCYSTLKAKGHNNMSYLTKSWCELCKAFLQEAKWSNNKIVPTESK
Glyma20g18280_1_FG (336) NTVNDTAYSILKEKGFNNLSYLKKSWCELCKAFLQEAKWSNNKIVPAESK
P_montana (336) NTVNDTSYSILKEKGHNNLSYLTKSWRELCKAFLQEAKWSNNKITPAESK
C_limon (337) NFVNEFAYYVLKQQDFLMLLSTKNAWLCLIQAYLVEAKWYHSKYTPKLEE
Q_ilex (337) NSVNEMALDTMKEQRFHIIKYLAKAWVDLCEYYLVEAKWYSNKYRPSLQE
V_vinifera (337) NSTNEMAYDALKEHGLPIVSYLRKAWSDLCKSYLLEAKWYYSPYTPSLQE
C_tenuipile (334) NSINEIGYDILKEEGRMVIPYLRNTWTELCKAFLVEAKWYSSGYTPTLEE
E.globulus_trunc (342) NSVNEMAYETLKEKGEMVIPYLAKAWYDLCKAFLQEAKWSNSRIIPGVEE
M_alternifolia (342) NSVNDMAYETLKETGEMVIPYLTKVWYDLCKAFLQEAKWSYNKITPGVEE
Glyma06g45780.1 (314) NTVNEFAYDALKEQGQNILPYLTKAWSNMLKAFLEEAKWSRLKHVPKEDD
Glyma12g10990.1 (334) NTVNEFAYDALKIKGQNILPHLTKAWSVMLKAFLQEAKWCRLKYLPPEED
L_japonicus (334) NTVNELAYDALREQGHDILPYLTKAWSDMLKAFLQEAKWCREKHLPKEEH
M.truncatula (297) TTVNDFTYDTIKETGHDILPYLVKVWSDMLKAFLQEAKWCHNKHMPKEDD
Q_petraea (336) NTVNEMVYDTLKEQGVDILPYLTKAWGDICKAFLQETKWRYYKRTPSSED
P.alba (337) NTINEIAYDNLKDKGENILPYLTKAWADLCNAFLQEAKWLYNKSTPTEDD
S.alba (337) NTINEIAYDNLKEKGENILPYLTKAWADLCNAFLQEAKWLYNKSTPTESD
Consensus (351) NTVNEMAYD LKE G NILPYLTKAW DLCKAFLQEAKW NK IP FED

Figure 2E


401 450
A_hypogaea (389) YTRNASVSSSSGALLAPCYESVTQDST--S--CAIDSITNYHGIVRSSCA
Glyma09g21900.1 (353) YLENASVSSSGMALLTASYESVCQQQD-ISNQQALCSLTNEQGLVRSSSN
Glyma20g18280_1_PG (386) YLENASVSSSGVALLAPSYESVCQEQDISFSDKTLHYLTNEGGLVRSSCT
P_montana (386) YLENASVSSSGVALLAPSYESVCQQQE-DISDHALRSLTDEHGLVRSSCV
C_limon (387) YLENGLVSTTGPLIIAISYLSGTNPII----KKEIEFLESNPDIVHWSSK
Q_ilex (387) YIENAWISIGAPTILVHAYEPVTNPTT----KLALDCLEEYPNLIRWSSI
V_vinifera (387) YISNSWISISGPVILVHAYFLVANPIT----KEALCSLERYHNLIRWSSM
C_tenuipile (384) YLQTSWISIGSLPMQTYVEALLGKNLAP----ESSDFAEKISDLLRLGGM
E.globulus_trunc (392) YLNNGWVSSSGSVMLIHAYELASPSIR----KEELESLEHYHDLLRLPSL
M_alternifolia (392) YLNNGWVSSSGQVMLTHAYFLSSPSLR----KEELESLEHYHDLLRLPSL
Glyma06g45780.1 (364) YLNNAWVSVSGVVILTHAYELLNHSIT----KEALCSLENYHALLRRSST
Glyma12g10990.1 (384) YLNNAWVSVSGVVILTHAYELLNQNIT----KDALDSLDNYHDLLRRPSI
L_japonicus (384) YLNNAWVSVSGVVILTHAYELLNHNTT----KEVLEALENYHALLKRPSI
M.truncatula (347) YLNNAWVSVSGVVELTHSYELLNRNIT----KEGLGYLENCPMLLQTPSI
Q_petraea (386) YLDNAWISVSGALLLIHAYELMSPSIT----DFALKGLEDYHNILRWPSI
P.alba (387) YFGNAWKSSSGPLQLVFAYEAVVQNIK----KEEIENLQKYHDTISRPSH
S.alba (387) YFGNAWKSSSGPLQLVFAYEAVVQNIK----KEEIENLLKYHDLISWPSY
Consensus (401) YL NAWVSVSG VIL HAYELV Q I KEALESLE YH LLR SI
451 500
A_hypogaea (435) IFRLCNDLATSAAELRGETTNSITSYMTEN-GTIEEEARESLGKLIDQE
Glyma09g21900.1 (402) IFRLCNDLATSAAELETGETANSITCYMHEK-DTSEEQAREELTNLIDAE
Glyma20g18280_1_PG (436) IFRLCNDLITSAAELRGETTNSIMSYMHEN-GTSEEHACEELRNLIDIE
P_montana (435) IFRLCNDLATSAAELRGETTNSIISYMHENDGTSEEQAREELPKLIDAE
C_limon (433) IFRLQDDLGTSSDEIQRGDVPKSIQCYMHET-GASEEVAREHIKDMMRQM
Q_ilex (433) IARLADDLGTSTDELKRGDVPKAIQCYMNET-GASEEGAREYIKYLISAT
V_vinifera (433) ILRLSDDLGTSLDELKRGDVPKSIQCYMYET-GASEEDAEHTSYLIGET
C_tenuipile (430) MIRLPDDLGTSTDELKRGDVPKSIQCYMHEA-GVTEDVARDHIMGLFQET
E.globulus_trunc (438) IFRLTNDIASSSAELRGETTNSIRCFMQEK-GISELEARECVKEEIDTA
M_alternifolia (438) IFRLTNDLATSSAELGRGETTNSILCYMREK-GFSESEARKQVIEQIDTA
Glyma06g45780.1 (410) IFRLCNDLGTSAAELERGEAASSIVCYMRES-GASEEGAYKHIRRLINET
Glyma12g10990.1 (430) IFRLCNDLGTSRAELQRGEAASSIVCNMRES-CVTEEGAYKNIHSLLDET
L_japonicus (430) IFRLCNDLGTSTAELQRGEVANSILSCMHEN-DICEESAHQHIHSLINET
M.truncatula (393) IFRLCNDLATSSAELERGEGANSIICYMNEN-GVSEEVAYKHIQNLLDQT
Q_petraea (432) IFRLTNDLGTSTAELERGETANSILCYMRET-SRSEDFAREHISNLIDKT
P.alba (433) IFRLCNDLASASAEIARGETANSVSCYMRTK-GISEELATESVMNLIDET
S.alba (433) IFRLCNDLASASAEIARGETANSVSCYMRTK-GISEELATTSVMNLIDET
Consensus (451) IFRLCNDLATSSAEL RGET NSI CYM E G SEE ARE I LID T

Figure 2F

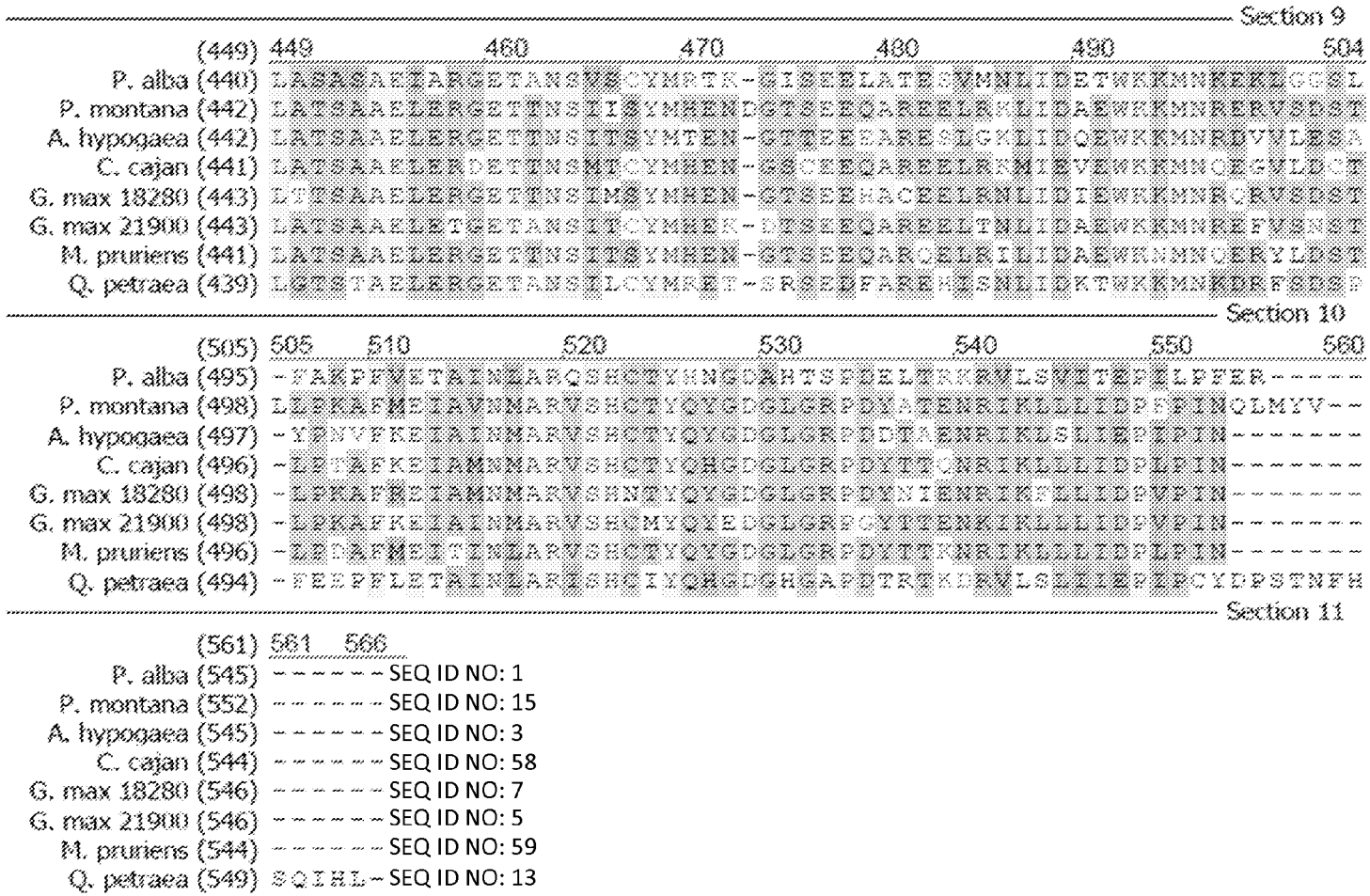
Figure 3A


Section 1
Section 2
Section 3
Section 4
P. alba
P. montana
A. hypogaea
C. cajan
G. max 18281
G. max 21940
M. pruriens
Q. petraea

Figure 3B


Section 5
Section 6
Section 7
Section 8
P. alba
P. montana
A. hypoqaea
C. cajan
G. max 18280
G. max 21900
M. pruriens
Q. petraea

Figure 3C


Section 9
(449) 449 460 470 480 490 504
P. alba (440)
P. montana (442)
A. hypogaea (442)
C. cajan (441)
G. max 18280 (443)
G. max 21900 (443)
M. pruriens (441)
Q. petraea (439)
Section 10
(505) 505 510 520 530 540 550 560
P. alba (495)
P. montana (498)
A. hypogaea (497)
C. cajan (496)
G. max 18280 (498)
G. max 21900 (498)
M. pruriens (496)
Q. petraea (494)
Section 11
(561) 561 566
P. alba (545) ------ SEQ ID NO: 1
P. montana (552) ------ SEQ ID NO: 15
A. hypogaea (545) ------ SEQ ID NO: 3
C. cajan (544) ------ SEQ ID NO: 58
G. max 18280 (546) ------ SEQ ID NO: 7
G. max 21900 (546) ------ SEQ ID NO: 5
M. pruriens (544) ------ SEQ ID NO: 59
Q. petraea (549) SQIHL- SEQ ID NO: 13

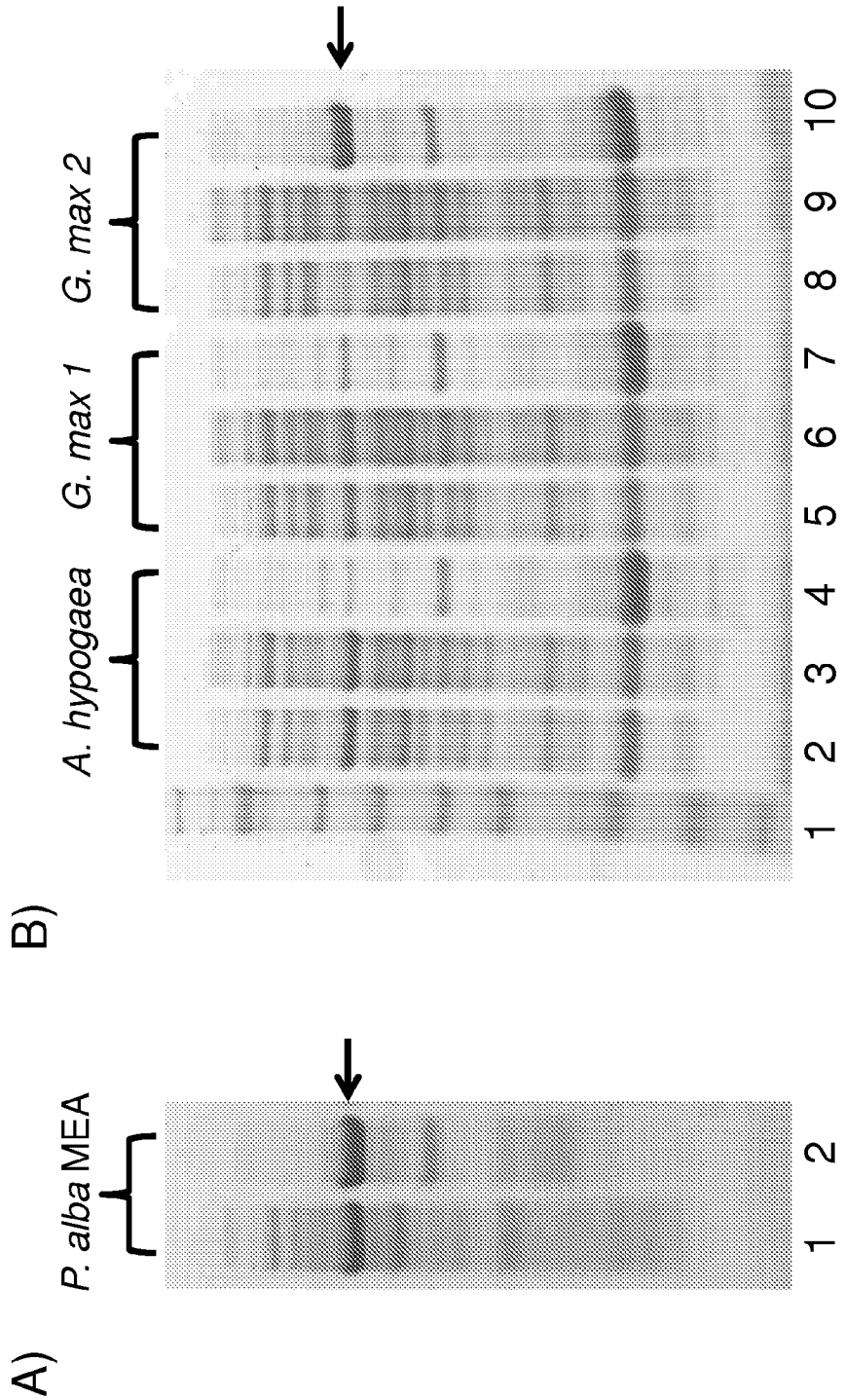



Figure 4

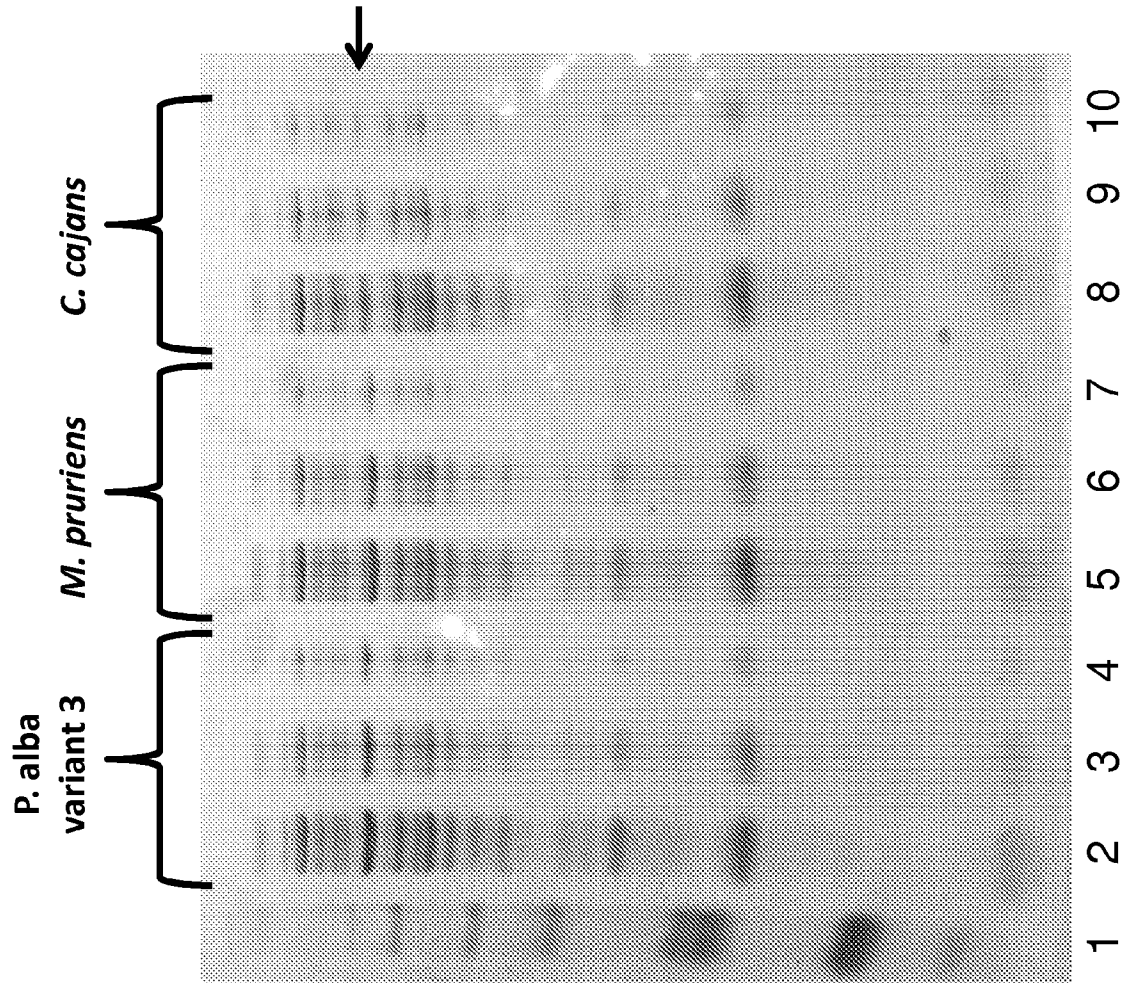



Figure 5

Figure 6

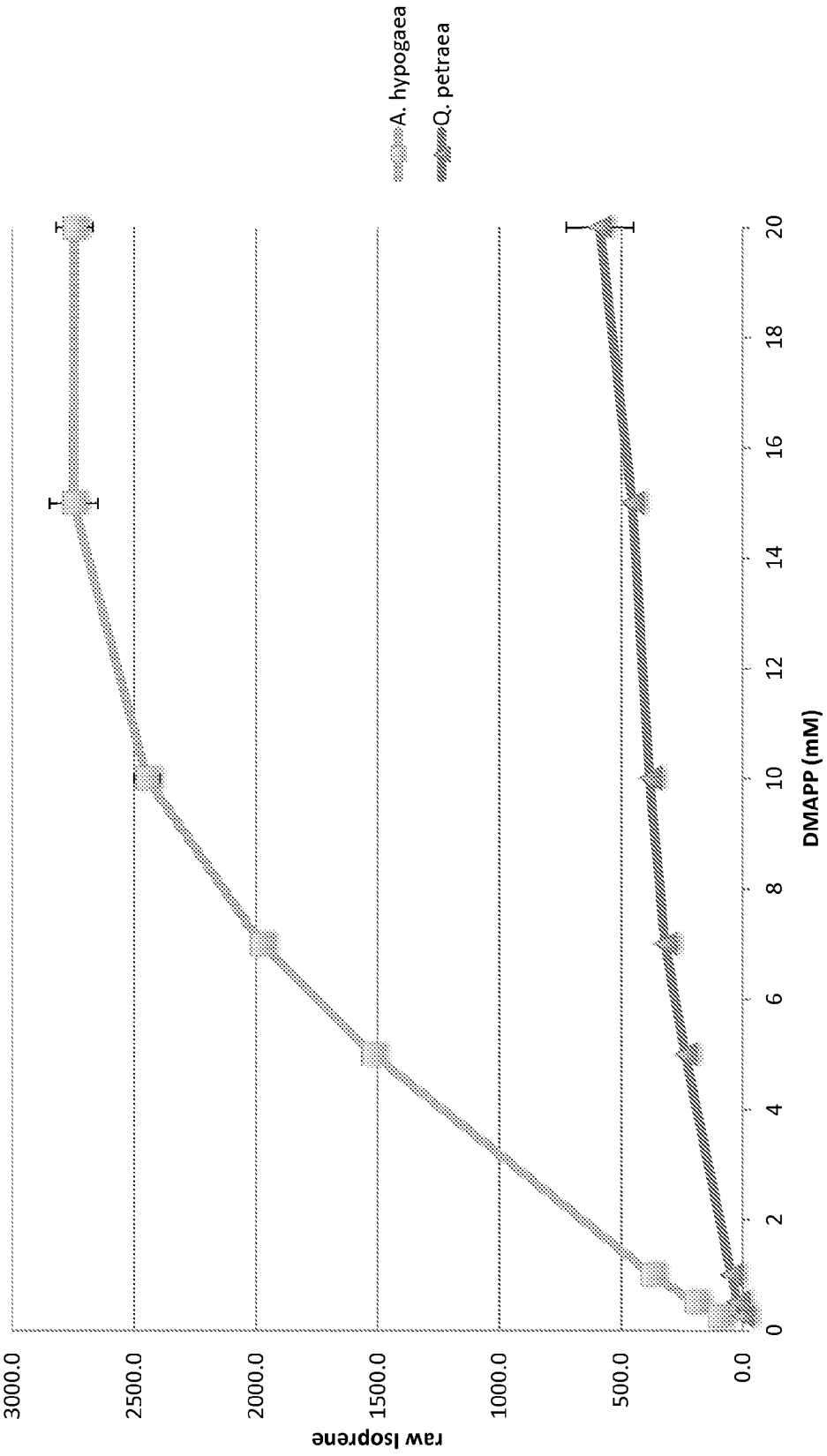

A. hypogaea
Q. petraea
raw Isoprene
3000.0
2500.0
2000.0
1500.0
1000.0
500.0
0.0
0
2
4
6
8
10
12
14
16
18
20
DMAPP (mM)

Figure 7

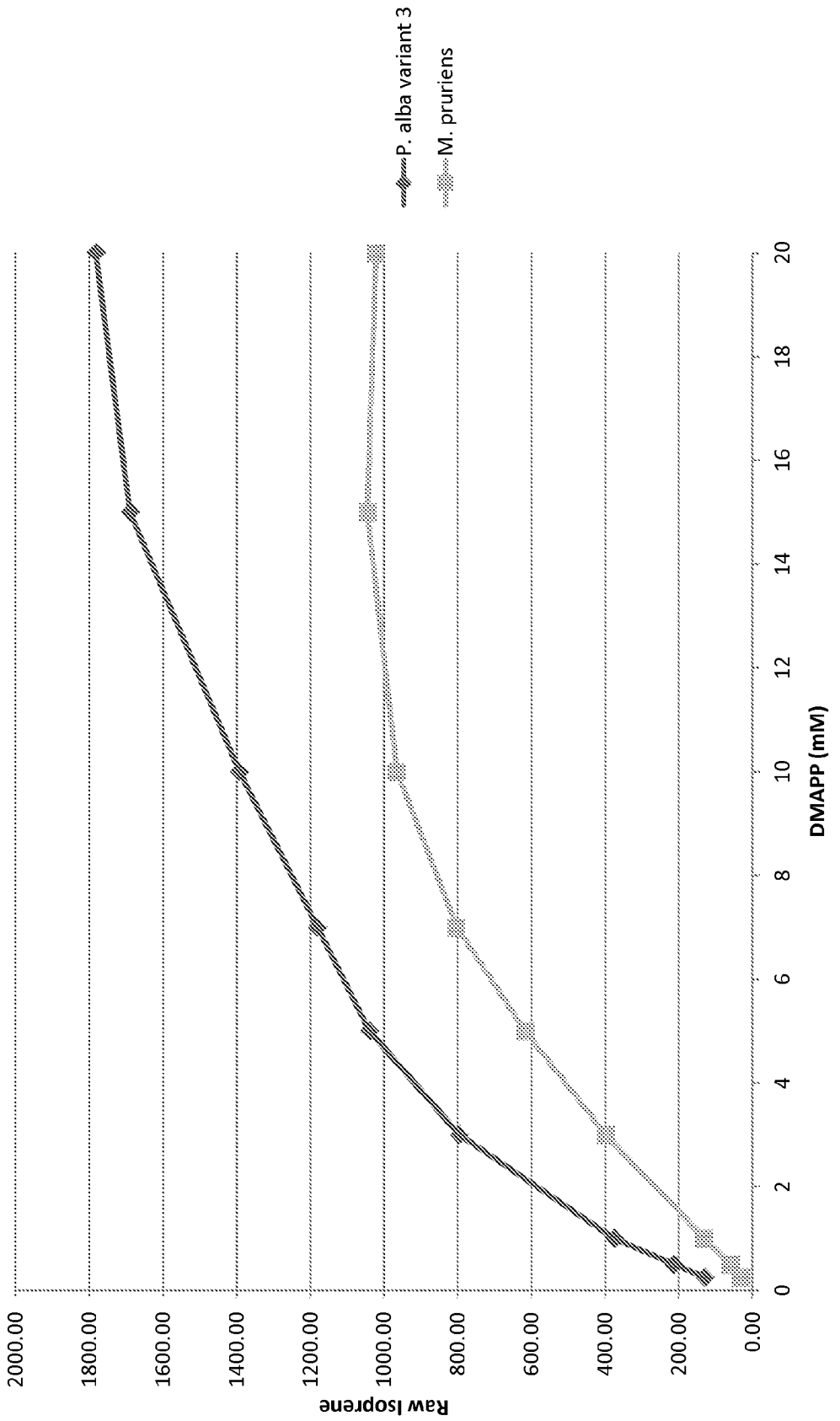

P. alba variant 3
M. pruriens
Raw Isoprene
2000.00
1800.00
1600.00
1400.00
1200.00
1000.00
800.00
600.00
400.00
200.00
0.00
0
2
4
6
8
10
12
14
16
18
20
DMAPP (mM)

Figure 8

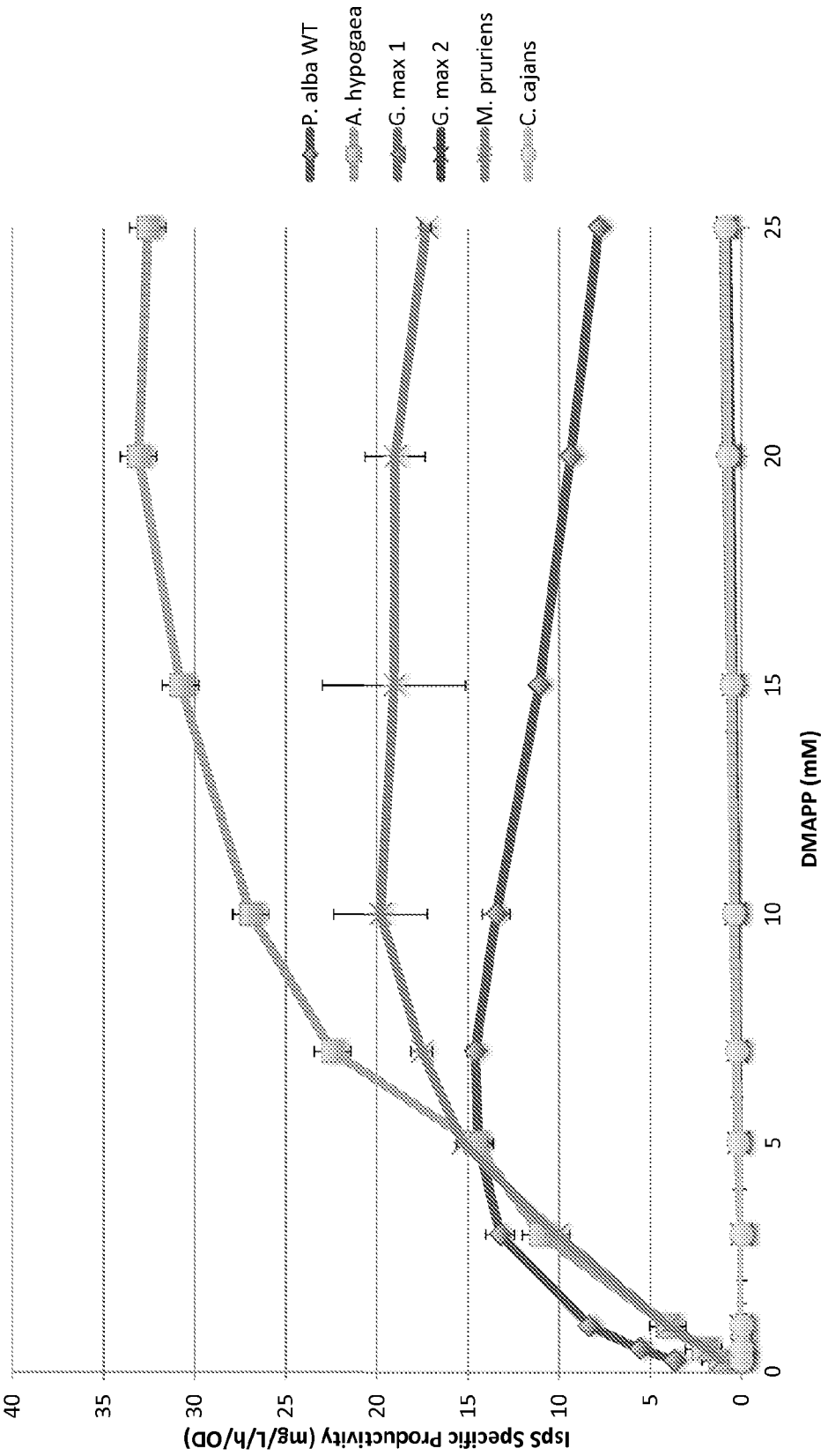

P. alba WT
A. hypogaea
G. max 1
G. max 2
M. pruriens
C. cajans
Isps Specific Productivity (mg/L/h/OD)
DMAPP (mM)
0
5
10
15
20
25
30
35
40

Figure 9

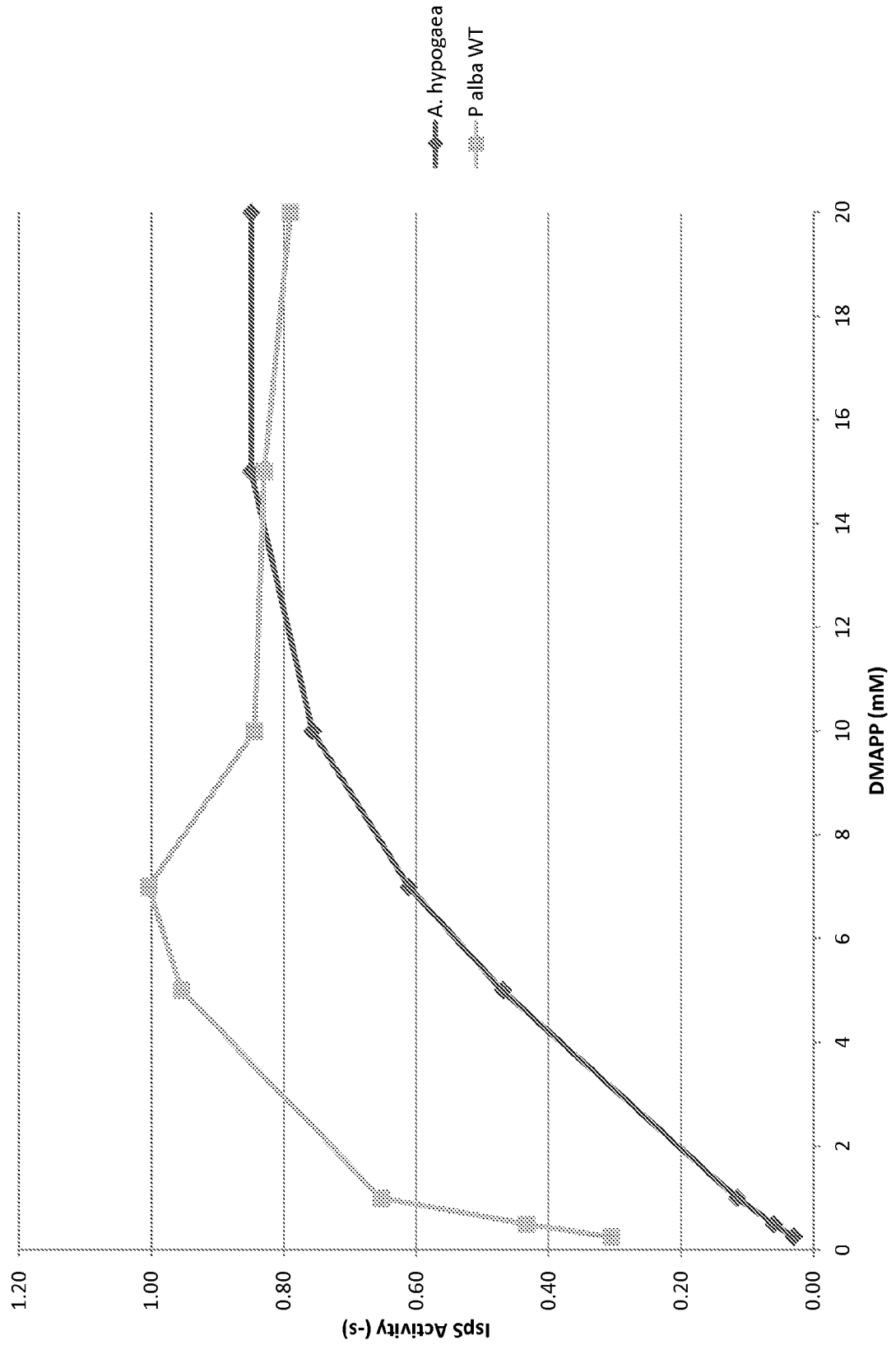

A. hypogaea
P alba WT
IspS Activity (-s)
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
2
4
6
8
10
12
14
16
18
20
DMAPP (mM)

Figure 10

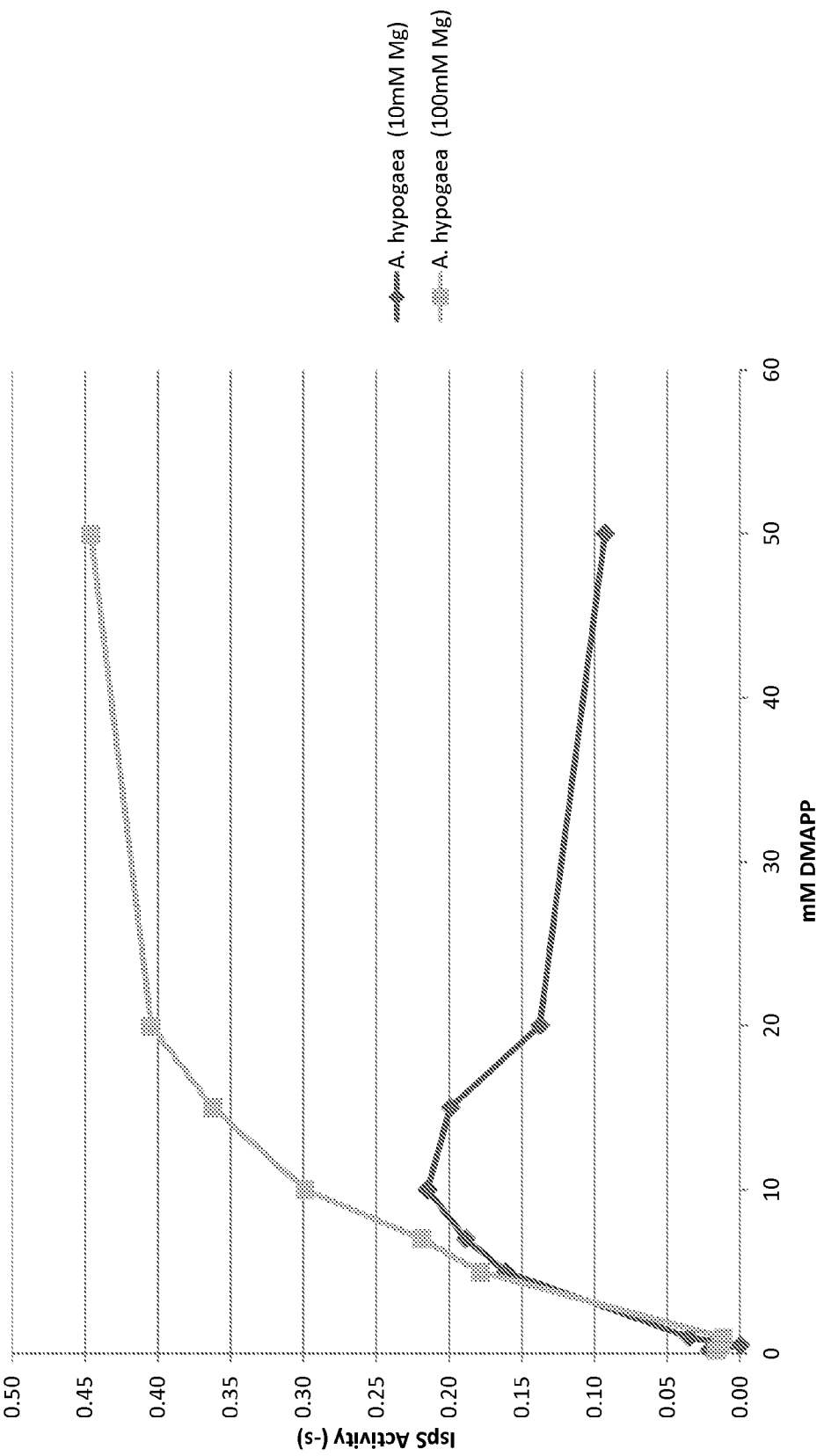
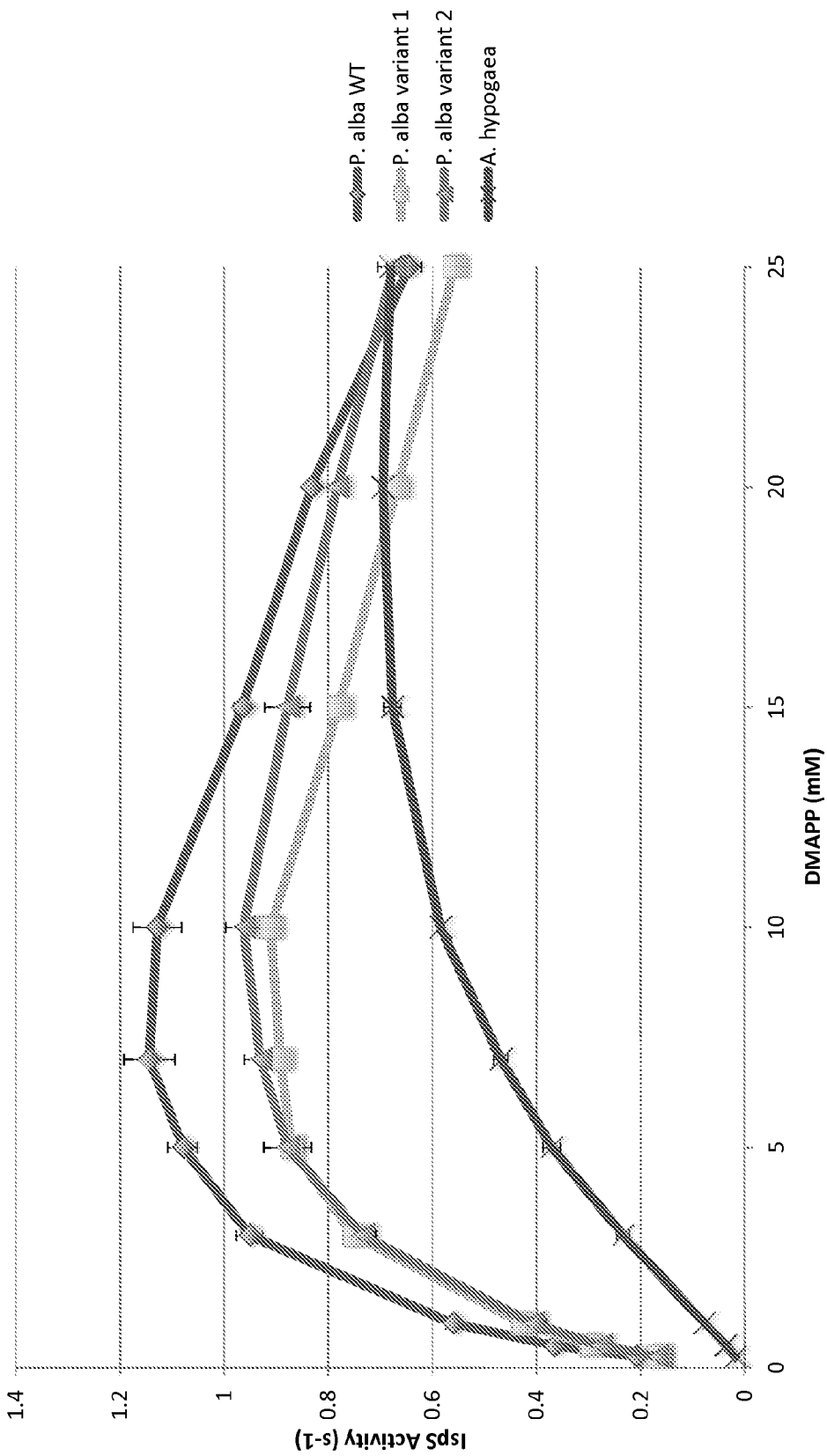

A. hypogaea (10mM Mg)
A. hypogaea (100mM Mg)
IspS Activity (-s)
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
mM DMAPP

Figure 11


P. alba WT
P. alba variant 1
P. alba variant 2
A. hypogaea
IspS Activity (s-1)
DMAPP (mM)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
5
10
15
20
25

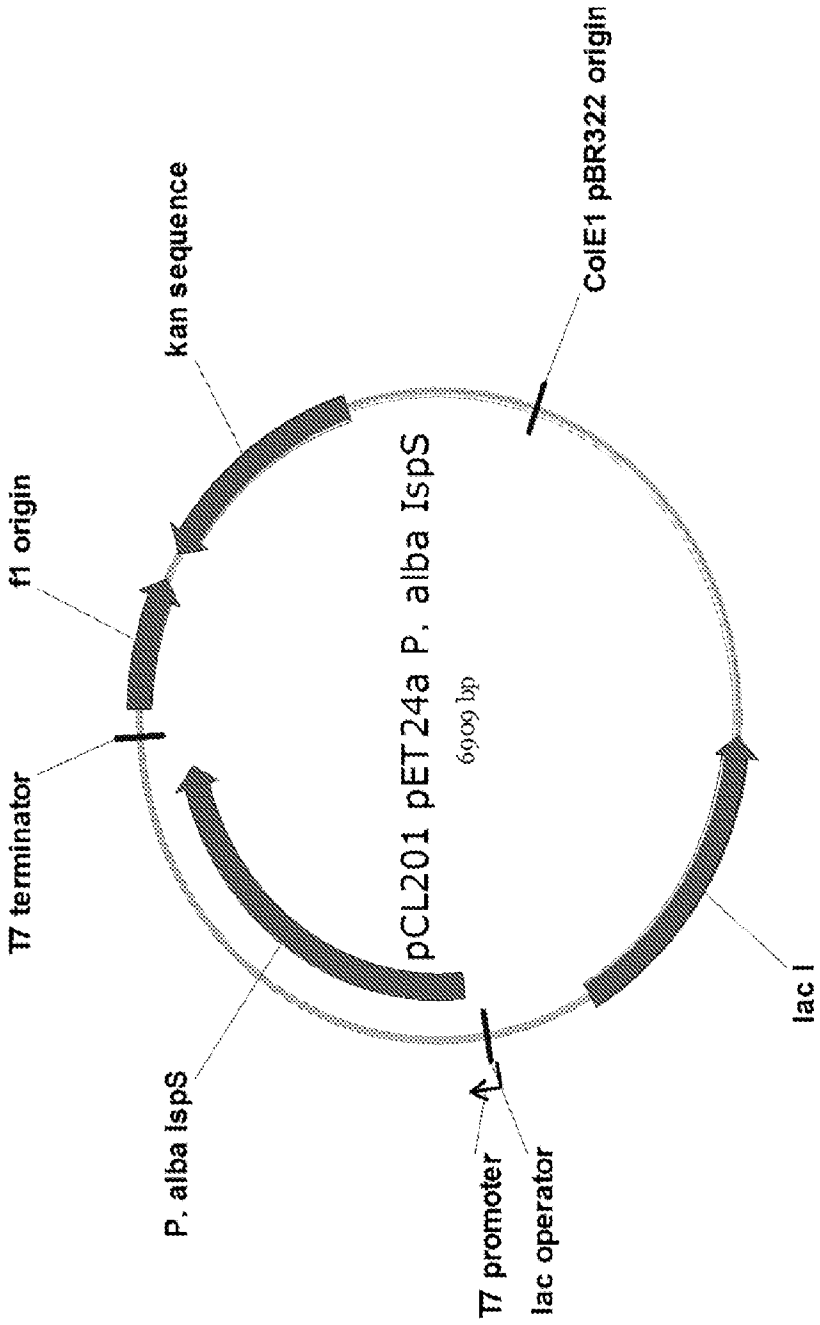

pCL201 pET24a P. alba IspS
6909 bp
kan sequence
ColE1 pBR322 origin
f1 origin
T7 terminator
P. alba ispS
T7 promoter
lac operator
lac I

Figure 12

Figure 13

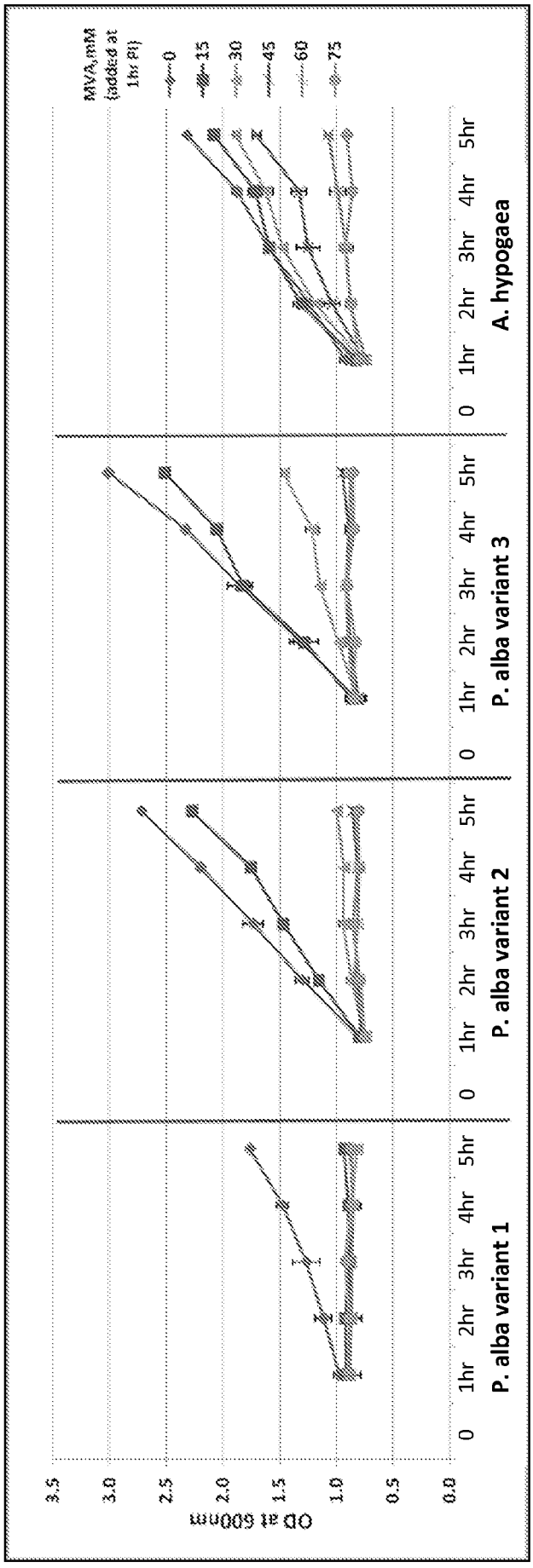
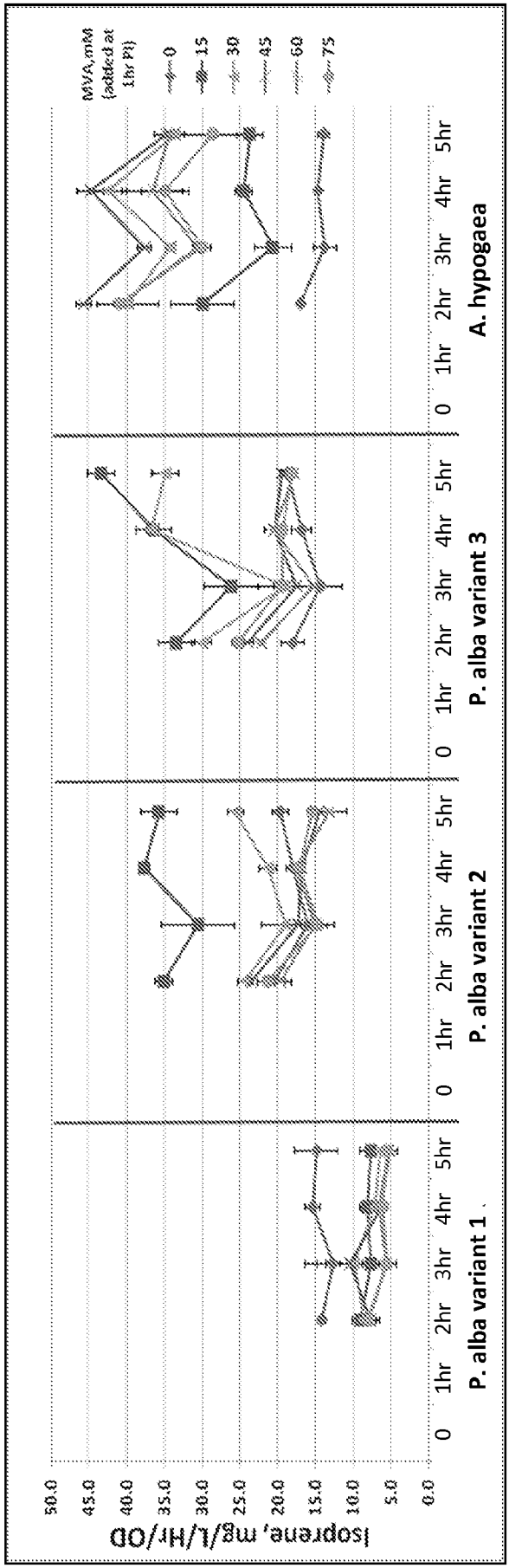

MVA,mM (added at 1hr pi)
0
15
30
45
60
75
OD at 600nm
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0 1hr 2hr 3hr 4hr 5hr
P. alba variant 1
P. alba variant 2
P. alba variant 3
A. hypogaea

Figure 14


Isoprene, mg/L/Hr/OD
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
0 1hr 2hr 3hr 4hr 5hr
P. alba variant 1
P. alba variant 2
P. alba variant 3
A. hypogaea
MVA, mM (added at 1hr PI)
0
15
30
45
60
75

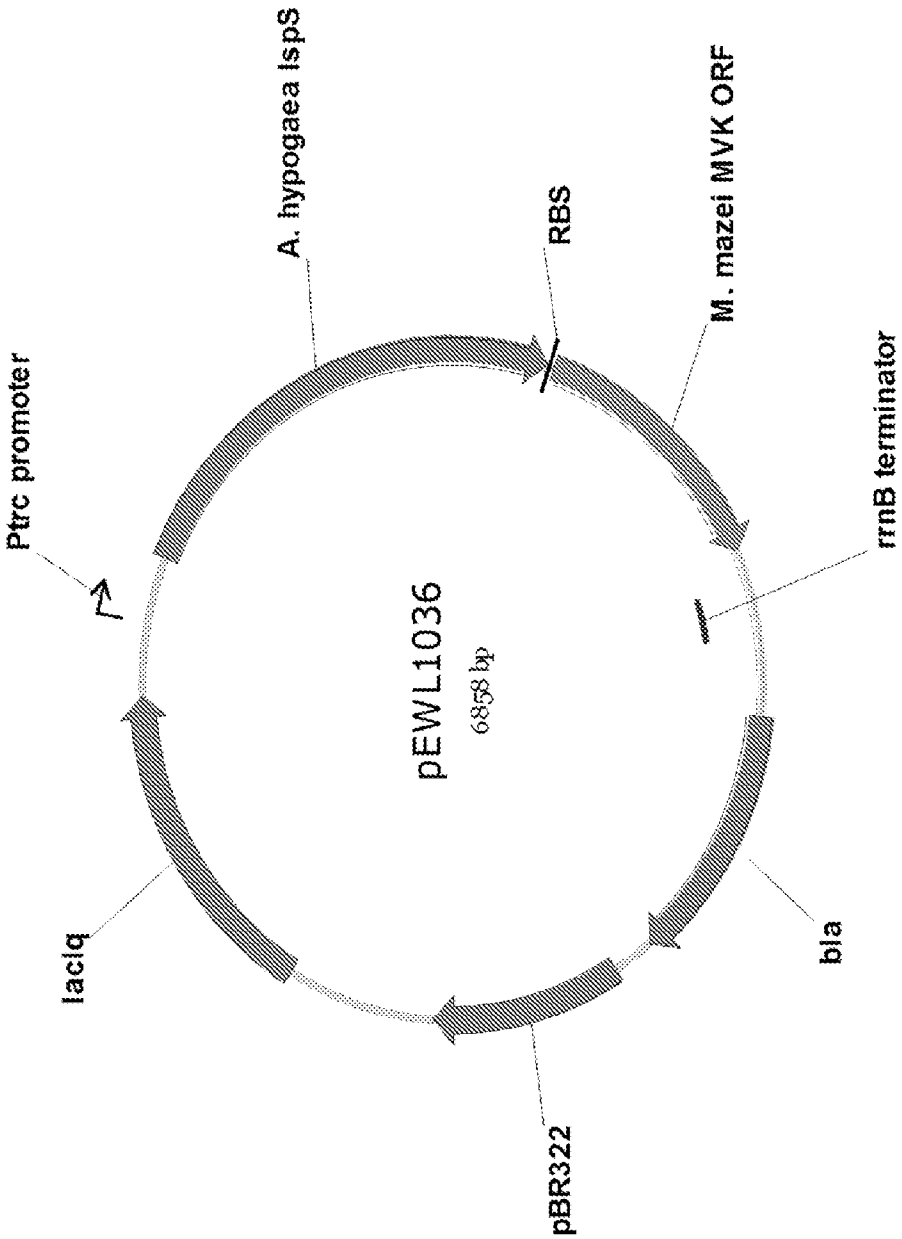

A. hypogaea IspS
RBS
M. mazei MVK ORF
rrnB terminator
bla
pBR322
lacIq
Ptrc promoter
pEWL1036
6858 bp

Figure 15

Figure 16

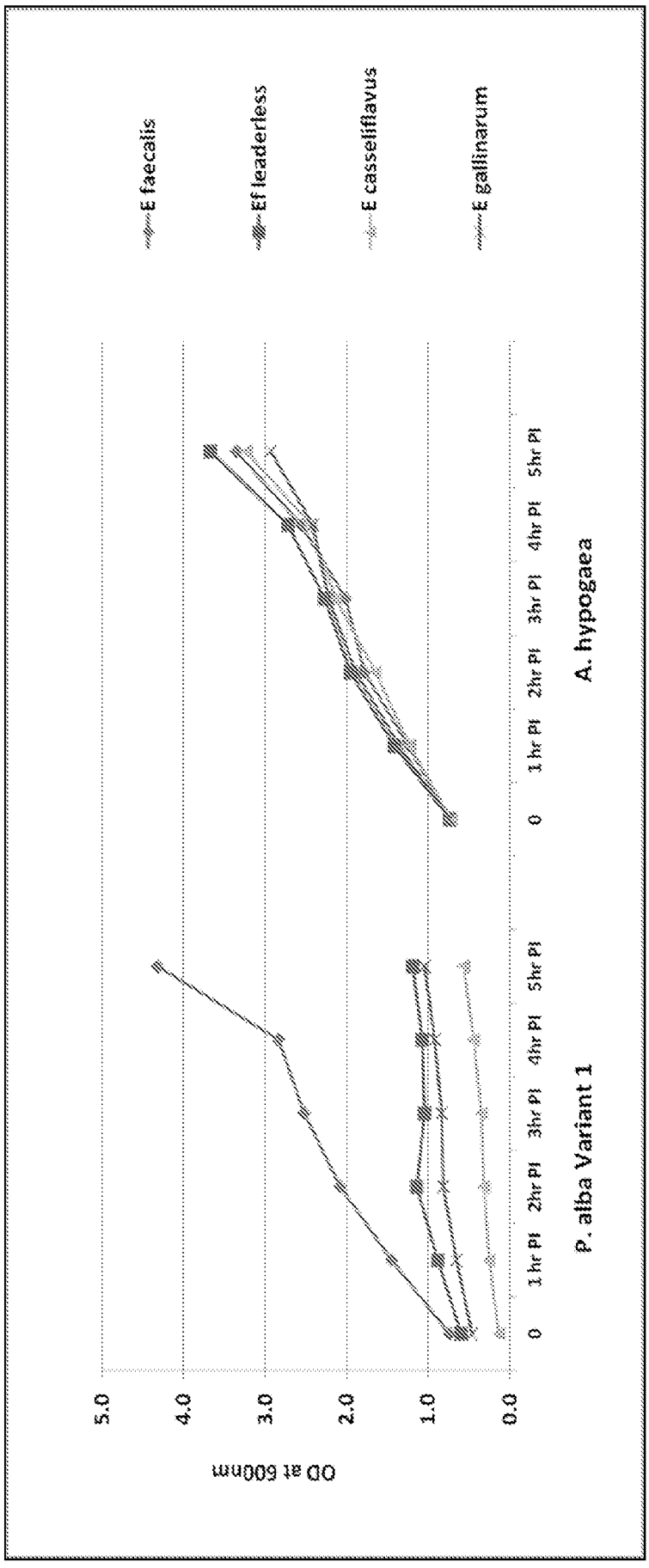
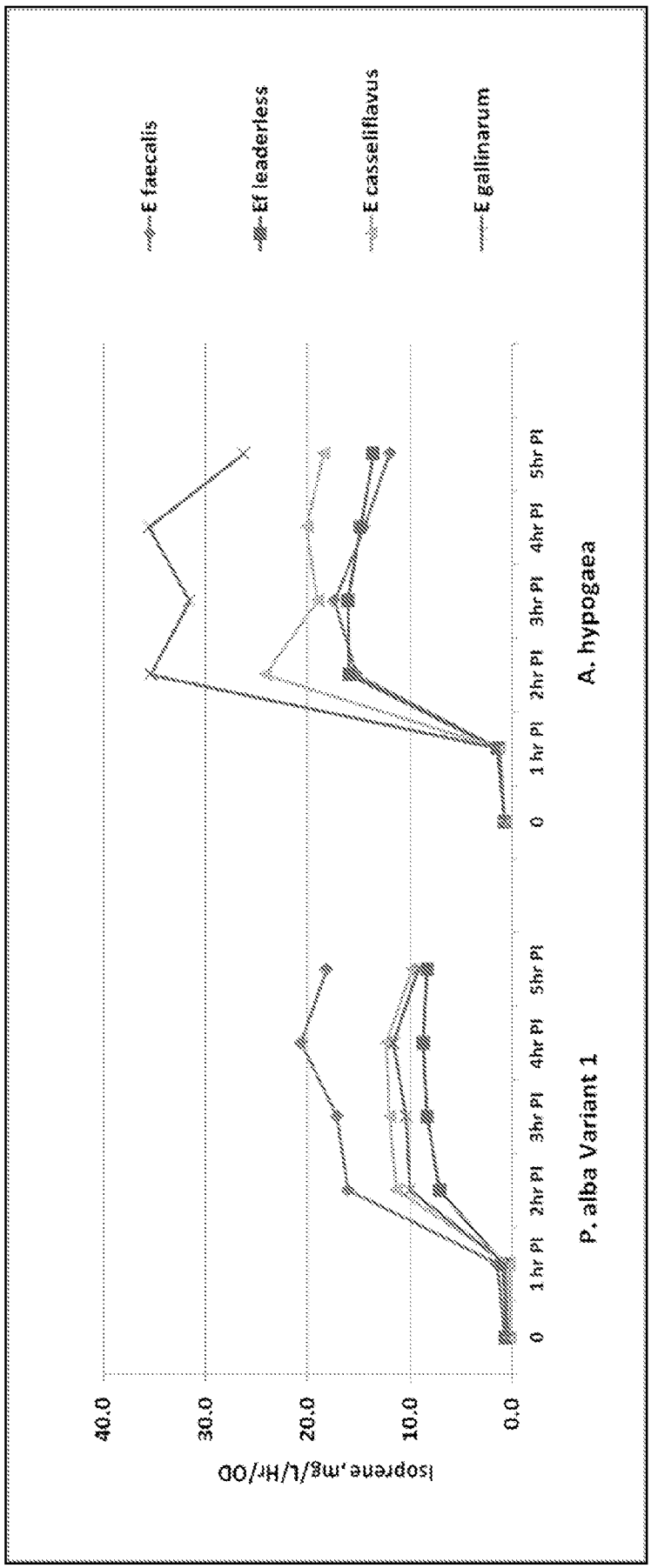

E faecalis
Ef leaderless
E casseliflavus
E gallinarum
OD at 600nm
5.0
4.0
3.0
2.0
1.0
0.0
0
1 hr PI
2hr PI
3hr PI
4hr PI
5hr PI
P. alba Variant 1
A. hypogaea

Figure 17


E faecalis
Ef leaderless
E casseliflavus
E gallinarum
Isoprene,mg/L/Hr/OD
40.0
30.0
20.0
10.0
0.0
0
1 hr PI
2hr PI
3hr PI
4hr PI
5hr PI
P. alba Variant 1
A. hypogaea

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/039315

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/415 C12N9/88
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)
C07K C12N

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 practicable, search terms used)
EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, EMBASE, PAJ, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2009/132220 A2 (DANISCO US INC [US]; GOODYEAR TIRE & RUBBER [US]; CERVIN MARGUERITE A) 29 October 2009 (2009-10-29) claims 1-53; sequences 122, 152 | 1-46 |
| X | DATABASE UniProt [Online]<br>21 March 2012 (2012-03-21), "SubName: Full=Isoprene synthase; Flags: Fragment;", XP002702769, retrieved from EBI accession no. UNIPROT:H2CSU6 Database accession no. H2CSU6 the whole document | 1-11, 18-23 |

-/--

[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 application or patent 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 |
|---|---|
| 16 July 2013 | 29/07/2013 |

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

# INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/039315

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | DATABASE UniProt [Online]<br><br>21 March 2012 (2012-03-21),<br>"SubName: Full=Isoprene synthase; Flags: Fragment;",<br>XP002702770,<br>retrieved from EBI accession no. UNIPROT:H2CSU7<br>Database accession no. H2CSU7<br>the whole document | 1-11, 18-23 |
| Y | JIANMING YANG ET AL: "Bio-isoprene production using exogenous MVA pathway and isoprene synthase in",<br>BIORESOURCE TECHNOLOGY, ELSEVIER BV, GB,<br>vol. 104, 12 October 2011 (2011-10-12),<br>pages 642-647, XP028351244,<br>ISSN: 0960-8524, DOI: 10.1016/J.BIORTECH.2011.10.042<br>[retrieved on 2011-10-20]<br>the whole document | 1-46 |
| Y | WO 2010/148150 A1 (DANISCO US INC [US]; GOODYEAR TIRE & RUBBER [US]; BECK ZACHARY QUINN [] 23 December 2010 (2010-12-23)<br>claims 1-8 | 1-46 |
| Y | MILLER BARBARA ET AL: "First isolation of an isoprene synthase gene from poplar and successful expression of the gene in Escherichia coli",<br>PLANTA, SPRINGER VERLAG, DE,<br>vol. 213, no. 3, 1 July 2001 (2001-07-01),<br>pages 483-487, XP009107808,<br>ISSN: 0032-0935, DOI: 10.1007/S004250100557<br>the whole document | 1-46 |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2013/039315

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2009132220 A2 | 29-10-2009 | AU 2009240505 A1 | 29-10-2009 |
| | | CA 2722440 A1 | 29-10-2009 |
| | | CN 102171335 A | 31-08-2011 |
| | | EP 2279251 A2 | 02-02-2011 |
| | | JP 2011518564 A | 30-06-2011 |
| | | KR 20110020234 A | 02-03-2011 |
| | | RU 2010147659 A | 27-05-2012 |
| | | US 2010003716 A1 | 07-01-2010 |
| | | WO 2009132220 A2 | 29-10-2009 |
| WO 2010148150 A1 | 23-12-2010 | AR 077115 A1 | 03-08-2011 |
| | | CA 2765805 A1 | 23-12-2010 |
| | | CN 102753697 A | 24-10-2012 |
| | | EP 2443244 A1 | 25-04-2012 |
| | | SG 176936 A1 | 28-02-2012 |
| | | TW 201111512 A | 01-04-2011 |
| | | US 2011014672 A1 | 20-01-2011 |
| | | WO 2010148150 A1 | 23-12-2010 |