

(19)



(11) Publication number:

SG 172646 A1

(43) Publication date:

28.07.2011

(51) Int. Cl:

;

(12)

Patent Application

(21) Application number: **2011038882**

(71) Applicant:

**AMYRIS, INC. 5980 HORTON STREET,
SUITE 450, EMERYVILLE, CA 94608 CA
US**

(22) Date of filing: **25.05.2007**

(30) Priority: **US 60/808,989 26.05.2006**

US 60/870,592 18.12.2006

(72) Inventor:

**RENNINGER, NEIL, STEPHEN 4058
LYMAN ROAD, OAKLAND, CA 94602
CALIFORNIA UNITED STATES OF
AMERICA US
NEWMAN, JACK 1715 SACRAMENTO
STREET, BERKELEY, CA 94702
CALIFORNIA UNITED STATES OF
AMERICA US
REILING, KEITH, KINKEAD 3534
RHODA AVENUE, OAKLAND, CA 94602
CALIFORNIA UNITED STATES OF
AMERICA US
REGENTIN, RIKI 25409 MODOC
COURT, HAYWARD, CA 94542
CALIFORNIA UNITED STATES OF
AMERICA US
PADDON, CHRISTOPHER, JOHN 1046
BANYAN WAY, PACIFICA, CA 94044
CALIFORNIA UNITED STATES OF
AMERICA US**

(54) Title:

PRODUCTION OF ISOPRENOIDS

(57) Abstract:

PRODUCTION OF ISOPRENOIDS ABSTRACT The present invention provides methods for a robust production of isoprenoids via one or more biosynthetic pathways. The invention also provides nucleic acids, enzymes, expression vectors, and genetically modified host cells for carrying out the subject methods. The invention also provides fermentation methods for high productivity of isoprenoids from genetically modified host cells. (No Suitable Figure)

PRODUCTION OF ISOPRENOIDS

ABSTRACT

The present invention provides methods for a robust production of isoprenoids via one or more biosynthetic pathways. The invention also provides nucleic acids, enzymes, expression vectors, and genetically modified host cells for carrying out the subject methods. The invention also provides fermentation methods for high productivity of isoprenoids from genetically modified host cells.

(No Suitable Figure)

PRODUCTION OF ISOPRENOIDS

CROSS-REFERENCE

- [0001] This application claims priority to U.S. Provisional Application Nos. 60/808,989, filed May 26, 2006 and 60/870,592 filed on December 18, 2006, which are incorporated herein by reference in their entirety.

BACKGROUND OF THE INVENTION

- [0002] Isoprenoids are ubiquitous in nature. They comprise a diverse family of over 40,000 individual products, many of which are vital to living organisms. Isoprenoids serve to maintain cellular fluidity, electron transport, and other metabolic functions. A vast number of natural and synthetic isoprenoids are useful as pharmaceuticals, cosmetics, perfumes, pigments and colorants, fungicides, antiseptics, nutraceuticals, and fine chemical intermediates.
- [0003] An isoprenoid product is typically composed of repeating five carbon isopentenyl diphosphate (IPP) units, although irregular isoprenoids and polyterpenes have been reported. In nature, isoprenoids are synthesized by consecutive condensations of their precursor IPP and its isomer dimethylallyl pyrophosphate (DMAPP). Two pathways for these precursors are known. Eukaryotes, with the exception of plants, generally use the mevalonate-dependent (MEV) pathway to convert acetyl coenzyme A (acetyl-CoA) to IPP, which is subsequently isomerized to DMAPP. Prokaryotes, with some exceptions, typically employ only the mevalonate-independent or deoxyxylulose-5-phosphate (DXP) pathway to produce IPP and DMAPP. Plants use both the MEV pathway and the DXP pathway. See Rohmer et al. (1993) *Biochem. J.* 295:517-524; Lange et al. (2000) *Proc. Natl. Acad. Sci. USA* 97(24):13172-13177; Rohdich et al. (2002) *Proc. Natl. Acad. Sci. USA* 99:1158-1163.
- [0004] Traditionally, isoprenoids have been manufactured by extraction from natural sources such as plants, microbes, and animals. However, the yield by way of extraction is usually very low due to a number of profound limitations. First, most isoprenoids accumulate in nature in only small amounts. Second, the source organisms in general are not amenable to the large-scale cultivation that is necessary to produce commercially viable quantities of a desired isoprenoid. Third, the requirement of certain toxic solvents for isoprenoid extraction necessitates special handling and disposal procedures, and thus complicating the commercial production of isoprenoids.
- [0005] The elucidation of the MEV and DXP metabolic pathways has made biosynthetic production of isoprenoids feasible. For instance, microbes have been engineered to overexpress a part of or the entire mevalonate pathway for production of an isoprenoid named amorpho-4,11-diene (U.S. Patent Nos. 7,172,886 and 7,192,751). Other efforts have focused on balancing the pool of glyceraldehyde-3-phosphate and pyruvate, or on increasing the expression of 1-deoxy-D-xylulose-5-phosphate synthase (dxs) and IPP isomerase (idi). See Farmer et al. (2001) *Biotechnol. Prog.* 17:57-61; Kajiwarra et al. (1997) *Biochem. J.* 324:421-426; and Kim et al. (2001) *Biotechnol. Bioeng.* 72:408-415.
- [0006] Nevertheless, given the very large quantities of isoprenoid products needed for many commercial applications, there remains a need for expression systems and fermentation procedures that produce even more isoprenoids than available with current technologies. Optimal redirection of microbial metabolism toward isoprenoid production requires that the introduced biosynthetic pathway is properly engineered both to funnel carbon to isoprenoid production efficiently and to prevent build up of toxic levels of metabolic

intermediates over a sustained period of time. The present invention addresses this need and provides related advantages as well.

SUMMARY OF THE INVENTION

- [0007] The present invention provides compositions and methods for a robust production of isoprenoids by the use of isopentenyl pyrophosphate pathway enzymes that are under the control of at least one heterologous regulator or fermentation conditions, either alone or in combination. Non-limiting examples of suitable isoprenoids include: hemiterpenes (derived from 1 isoprene unit) such as isoprene; monoterpenes (derived from 2 isoprene units) such as myrcene; sesquiterpenes (derived from 3 isoprene units) such as amorphadiene; diterpenes (derived from four isoprene units) such as taxadiene; triterpenes (derived from 6 isoprene units) such as squalene; tetraterpenes (derived from 8 isoprenoids) such as β -carotene; and polyterpenes (derived from more than 8 isoprene units) such as polyisoprene.
- [0008] In one aspect, a method of producing an isoprenoid involves the steps of (a) obtaining a plurality of host cells that comprise an enzymatic pathway for making isopentenyl pyrophosphate wherein the all of the pathway enzymes are under control of at least one heterologous transcriptional regulator; and (b) culturing the host cells in a medium under conditions that are suboptimal as compared to conditions that would provide for a maximum specific growth rate for the host cells. In some embodiments, the pathway is the mevalonate pathway. In other embodiments, the pathway is the DXP pathway. In other embodiments, the at least one heterologous transcriptional regulatory sequence is inducible. In other embodiments, the pathway enzymes are under control of a single transcriptional regulator. In other embodiments, the pathway enzymes are under control of multiple heterologous transcriptional regulators.
- [0009] In some embodiments, the pathway comprises a nucleic acid sequence encoding a mevalonate pathway enzyme from a prokaryote having an endogenous mevalonate pathway. Exemplary prokaryotes having an endogenous mevalonate pathway include but are not limited to the genus *Enterococcus*, the genus *Pseudomonas*, and the genus *Staphylococcus*. In one embodiment, the mevalonate pathway enzyme is selected from acetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, and mevalonate kinase. In another embodiment, the heterologous nucleic acid sequence encodes a Class II HMG-CoA reductase.
- [0010] In another embodiment, the host cells are cultured in a medium wherein the nutrient and/or temperature level is maintained at a level below that which would provide for the maximum specific growth rate for the host cells. In another embodiment, the host cells are cultured in a medium where the carbon source is maintained at a level to provide for less than about 90%, 75%, 50%, 25%, 10%, or anywhere between 90% and 10% of the maximum specific growth rate. In another embodiment, the host cells are cultured in a medium where the nitrogen source is maintained at a level to provide for less than about 90%, 75%, 50%, 25%, 10%, or anywhere between 90% and 10% of the maximum specific growth rate. In another embodiment, the host cells are cultured in a medium where the temperature is maintained at a level to provide for less than about 90%, 75%, 50%, 25%, 10% or anywhere between 90% and 10% of the maximum specific growth rate. In another embodiment, the medium temperature is maintained at least about 2°C, 4°C, 5°C, 6°C, 8°C, 10°C, 15°C, or 20°C below the temperature that would provide for the maximum specific growth rate.
- [0011] In yet another embodiment, a method of producing an isoprenoid or isoprenoid precursor comprises the steps of (i) performing a fermentation reaction comprising a fermentation medium and a plurality of genetically modified host cells that produce the isoprenoid under conditions such that (a) the fermentation

medium is kept at a temperature lower than that which would provide for a maximum specific growth rate of said host cells; (b) the fermentation medium comprises a carbon source present in an amount that is lower than that which would provide for a maximum specific growth rate of the host cells; and/or (c) the fermentation medium comprises a nitrogen source present in an amount that is lower than that which would provide for a maximum specific growth rate of the host cells; (ii) recovering the isoprenoid produced under one or more conditions set forth in (a) through (c). In one aspect, the isoprenoid is produced under at least two of the conditions set forth in (a) through (c). In another aspect, the isoprenoid is produced under all of the conditions set forth in (a) through (c).

INCORPORATION BY REFERENCE

- [0012] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0013] Figure 1A is a schematic representation of the mevalonate ("MEV") pathway for the production of isopentenyl pyrophosphate ("IPP").
- [0014] Figure 1B is a schematic representation of the 1-deoxy-D-xylulose 5-diphosphate ("DXP") pathway for the production of isopentenyl pyrophosphate ("IPP") and dimethylallyl pyrophosphate ("DMAPP"). Dxs is 1-deoxy-D-xylulose-5-phosphate synthase; Dxr is 1-deoxy-D-xylulose-5-phosphate reductoisomerase (also known as IspC); IspD is 4-diphosphocytidyl-2C-methyl-D-erythritol synthase; IspE is 4-diphosphocytidyl-2C-methyl-D-erythritol synthase; IspF is 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; IspG is 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (IspG); and ispH is isopentenyl/dimethylallyl diphosphate synthase.
- [0015] Figure 2 is a schematic representation of the conversion of isopentenyl pyrophosphate ("IPP") and dimethylallyl pyrophosphate ("DMAPP") to geranyl pyrophosphate ("GPP"), farnesyl pyrophosphate ("FPP"), and geranylgeranyl pyrophosphate ("GGPP"), and the synthesis of various isoprenoids.
- [0016] Figure 3 shows a map of expression plasmid pMBIS-gpps.
- [0017] Figure 4 shows a map of expression plasmid pAM408.
- [0018] Figure 5 shows a map of expression plasmid pAM424.
- [0019] Figure 6 shows a map of expression plasmids pTrc99A-ADS, pTrc99A-FSA, pTrc99A-LLS, pTrc99A-LMS, pTrc99A-GTS, pTrc99A-APS, pTrc99A-BPS, pTrc99A-PHS, pTrc99A-TS, pTrc99A-CS, pTrc99A-SS, and pAM373.
- [0020] Figures 7A-C are schematics for the construction of plasmids pAM489-pAM498 and for pAM328.
- [0021] Figure 8 shows the higher specific activity and increased stability of the *Enterococcus faecalis* HMGR-CoA reductase (HMGR) compared to the *Saccharomyces cerevisiae* truncated HMG-CoA reductase (tHMGR).
- [0022] Figure 9 shows the relationship between dry cell weight ("DCW") per liter and OD₆₀₀.
- [0023] Figures 10A-B show the increased volumetric and specific amorpha-4,11-diene productivity of host strains carrying the *Staphylococcus aureus* HMGR and HMGS genes compared to host strains carrying the *Saccharomyces cerevisiae* tHMGR and HMGS genes.
- [0024] Figure 11A-B show the effect of lower temperature on amorpha-4,11-diene productivity of an *Escherichia coli* host strain.

- [0025] Figures 12A-D show the effect of reduced glucose levels on amorpha-4,11-diene productivity of an *Escherichia coli* host strain.
- [0026] Figures 13A-B show the combined effects of lower temperature and reduced glucose levels on amorpha-4,11-diene productivity of an *Escherichia coli* host strain.
- [0027] Figures 14A-E and 15A-E show the combined effects of lower temperature and reduced glucose and nitrogen levels on amorpha-4,11-diene productivity of an *Escherichia coli* host strain.
- [0028] Figure 16 shows production of amorpha-4,11-diene via the DXP pathway by an *Escherichia coli* host strain.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

- [0029] Unless defined otherwise, 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 belongs. Reference is made here to a number of terms that shall be defined to have the following meanings:
- [0030] The term “optional” or “optionally” means that the subsequently described feature or structure may or may not be present, or that the subsequently described event or circumstance may or may not occur, and that the description includes instances where a particular feature or structure is present and instances where the feature or structure is absent, or instances where the event or circumstance occurs and instances where the event or circumstance does not occur.
- [0031] The terms “metabolic pathway” is used herein to refer to a catabolic pathway or an anabolic pathway. Anabolic pathways involve constructing a larger molecule from smaller molecules, a process requiring energy. Catabolic pathways involve breaking down of larger molecules, often releasing energy.
- [0032] The term “mevalonate pathway” or “MEV pathway” is used herein to refer to the biosynthetic pathway that converts acetyl-CoA to IPP. The MEV pathway is illustrated schematically in Figure 1A.
- [0033] The term “deoxyxylulose 5-phosphate pathway” or “DXP pathway” is used herein to refer to the pathway that converts glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP. The DXP pathway is illustrated schematically in Figure 1B.
- [0034] The word “pyrophosphate” is used interchangeably herein with “diphosphate”.
- [0035] The terms “expression vector” or “vector” refer to a nucleic acid that transduces, transforms, or infects a host cell, thereby causing the cell to produce nucleic acids and/or proteins other than those that are native to the cell, or to express nucleic acids and/or proteins in a manner that is not native to the cell.
- [0036] The term “endogenous” refers to a substance or process that occurs naturally, e.g., in a non-recombinant host cell.
- [0037] The terms “enzymatic pathway for making isopentenyl pyrophosphate” refers to any pathway capable of producing isopentenyl pyrophosphate, including, without limitation, either the mevalonate pathway or the DXP pathway.
- [0038] The term “nucleic acid” refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically, or biochemically modified, non-natural, or derivatized nucleotide bases.

- [0039] The term “operon” is used to refer to two or more contiguous nucleotide sequences that each encode a gene product such as a RNA or a protein, and the expression of which are coordinately regulated by one or more controlling elements (for example, a promoter).
- [0040] The term “gene product” refers to RNA encoded by DNA (or vice versa) or protein that is encoded by an RNA or DNA, where a gene will typically comprise one or more nucleotide sequences that encode a protein, and may also include introns and other non-coding nucleotide sequences.
- [0041] The term “protein” refers to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.
- [0042] The term “heterologous nucleic acid” as used herein refers to a nucleic acid wherein at least one of the following is true: (a) the nucleic acid is foreign (“exogenous”) to (that is, not naturally found in) a given host cell; (b) the nucleic acid comprises a nucleotide sequence that is naturally found in (that is, is “endogenous to”) a given host cell, but the nucleotide sequence is produced in an unnatural (for example, greater than expected or greater than naturally found) amount in the cell; (c) the nucleic acid comprises a nucleotide sequence that differs in sequence from an endogenous nucleotide sequence, but the nucleotide sequence encodes the same protein (having the same or substantially the same amino acid sequence) and is produced in an unnatural (for example, greater than expected or greater than naturally found) amount in the cell; or (d) the nucleic acid comprises two or more nucleotide sequences that are not found in the same relationship to each other in nature (for example, the nucleic acid is recombinant).
- [0043] A “transgene” refers to a gene that is exogenously introduced into a host cell. It can comprise an endogenous or exogenous, or heterologous nucleic acid.
- [0044] The term “recombinant host” (also referred to as a “genetically modified host cell” or “genetically modified host microorganism”) denotes a host cell that comprises a heterologous nucleic acid of the invention.
- [0045] The term “exogenous nucleic acid” refers to a nucleic acid that is exogenously introduced into a host cell, and hence is not normally or naturally found in and/or produced by a given cell in nature.
- [0046] The term “regulatory element” refers to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate expression of a coding sequence and/or production of an encoded polypeptide in a host cell.
- [0047] The term “transformation” refers to a permanent or transient genetic change induced in a cell following introduction of new nucleic acid. Genetic change (“modification”) can be accomplished either by incorporation of the new DNA into the genome of the host cell, or by transient or stable maintenance of the new DNA as an episomal element. In eukaryotic cells, a permanent genetic change is generally achieved by introduction of the DNA into the genome of the cell. In prokaryotic cells, a permanent genetic change can be introduced into the chromosome or via extrachromosomal elements such as plasmids and expression vectors, which may contain one or more selectable markers to aid in their maintenance in the recombinant host cell.
- [0048] The term “operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a nucleotide sequence if the promoter affects the transcription or expression of the nucleotide sequence.

- [0049] The term "host cell" and "host microorganism" are used interchangeably herein to refer to any archae, bacterial, or eukaryotic living cell into which a heterologous nucleic acid can be or has been inserted. The term also relates to the progeny of the original cell, which may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to natural, accidental, or deliberate mutation.
- [0050] The term "synthetic" as used in reference to nucleic acids means the annealing of chemically synthesized oligonucleotide building blocks to form gene segments, which are then enzymatically assembled to construct the entire gene. Synthesis of nucleic acids via "chemical means" means that the component nucleotides were assembled *in vitro*.
- [0051] The term "natural" as applied to a nucleic acid, a cell, or an organism, refers to a nucleic acid, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in a non-pathological (un-diseased) organism that can be isolated from a source in nature and that has not been intentionally modified by a human in the laboratory is natural.
- [0052] The term "naturally occurring" as applied to a nucleic acid, an enzyme, a cell, or an organism, refers to a nucleic acid, enzyme, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism that can be isolated from a source in nature and that has not been intentionally modified by a human in the laboratory is naturally occurring.
- [0053] The term "biologically active fragment" as applied to a protein, polypeptide or enzyme refers to functional portion(s) of the proteins or polypeptide or enzyme. Functionally equivalents may have variant amino acid sequences may arise, e.g., as a consequence of codon redundancy and functional equivalency which are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Functionally equivalent proteins or peptides may alternatively be constructed via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged.
- [0054] The terms "isoprenoid", "isoprenoid compound", "isoprenoid product", "terpene", "terpene compound", "terpenoid", and "terpenoid compound" are used interchangeably herein. They refer to compounds that are capable of being derived from IPP.
- [0055] The singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an expression vector" includes a single expression vector as well as a plurality of expression vectors, and reference to "the host cell" includes reference to one or more host cells, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.
- [0056] Unless otherwise indicated, this invention is not limited to particular sequences, expression vectors, enzymes, host microorganisms, or processes, as such may vary in accordance with the understanding of those of ordinary skill in the arts to which this invention pertains in view of the teaching herein. Terminology used herein is for purposes of describing particular embodiments only and is not intended to be limiting.

Host Cells

- [0057] Any suitable host cell may be used in the practice of the present invention. In one embodiment, the host cell is a genetically modified host microorganism in which nucleic acid molecules have been inserted,

deleted or modified (*i.e.*, mutated; e.g., by insertion, deletion, substitution, and/or inversion of nucleotides), to either produce the desired isoprenoid compound or isoprenoid derivative, or effect an increased yield of the desired isoprenoid compound or isoprenoid derivative. In another embodiment, the host cell is capable of being grown in liquid growth medium. In contrast, a "control cell" is an alternative subject or sample used in an experiment for comparison purpose, and is typically a parental cell that does not contain the modification(s) made to a corresponding host cell.

- [0058] Illustrative examples of suitable host cells include any archae, prokaryotic, or eukaryotic cell. Examples of an archae cell include, but are not limited to those belonging to the genera: *Aeropyrum*, *Archaeoglobus*, *Halobacterium*, *Methanococcus*, *Methanobacterium*, *Pyrococcus*, *Sulfolobus*, and *Thermoplasma*. Illustrative examples of archae strains include but are not limited to: *Aeropyrum pernix*, *Archaeoglobus fulgidus*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Pyrococcus abyssi*, *Pyrococcus horikoshii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*.
- [0059] Examples of a procaryotic cell include, but are not limited to those belonging to the genera: *Agrobacterium*, *Alicyclobacillus*, *Anabaena*, *Anacystis*, *Arthrobacter*, *Azobacter*, *Bacillus*, *Brevibacterium*, *Chromatium*, *Clostridium*, *Corynebacterium*, *Enterobacter*, *Erwinia*, *Escherichia*, *Lactobacillus*, *Lactococcus*, *Mesorhizobium*, *Methylobacterium*, *Microbacterium*, *Phormidium*, *Pseudomonas*, *Rhodobacter*, *Rhodopseudomonas*, *Rhodospirillum*, *Rhodococcus*, *Salmonella*, *Scenedesmun*, *Serratia*, *Shigella*, *Staphylococcus*, *Streptomyces*, *Synnecoccus*, and *Zymomonas*.
- [0060] Illustrative examples of prokaryotic bacterial strains include but are not limited to: *Bacillus subtilis*, *Bacillus amyloliquefacines*, *Brevibacterium ammoniagenes*, *Brevibacterium immariophilum*, *Clostridium beigerinckii*, *Enterobacter sakazakii*, *Escherichia coli*, *Lactococcus lactis*, *Mesorhizobium loti*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Pseudomonas pudica*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, *Salmonella enterica*, *Salmonella typhi*, *Salmonella typhimurium*, *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Staphylococcus aureus*, and the like.
- [0061] In general, if a bacterial host cell is used, a non-pathogenic strain is preferred. Illustrative examples of non-pathogenic strains include but are not limited to: *Bacillus subtilis*, *Escherichia coli*, *Lactibacillus acidophilus*, *Lactobacillus helveticus*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Pseudomonas pudita*, *Rhodobacter sphaeroides*, *Rodobacter capsulatus*, *Rhodospirillum rubrum*, and the like.
- [0062] Examples of eukaryotic cells include but are not limited to fungal cells. Examples of fungal cell include, but are not limited to those belonging to the genera: *Aspergillus*, *Candida*, *Chrysosporium*, *Cryptococcus*, *Fusarium*, *Kluyveromyces*, *Neotrophodinium*, *Neurospora*, *Penicillium*, *Pichia*, *Saccharomyces*, *Trichoderma* and *Xanthophyllomyces* (formerly *Phaffia*).
- [0063] Illustrative examples of eukaryotic strains include but are not limited to: *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Candida albicans*, *Chrysosporium lucknowense*, *Fusarium graminearum*, *Fusarium venenatum*, *Kluyveromyces lactis*, *Neurospora crassa*, *Pichia angusta*, *Pichia finlandica*, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia opuntiae*, *Pichia pastoris*, *Pichia pijperi*, *Pichia quercuum*, *Pichia salictaria*, *Pichia thermotolerans*, *Pichia trehalophila*, *Pichia stipitis*, *Streptomyces ambofaciens*, *Streptomyces aureofaciens*, *Streptomyces aureus*, *Saccaromyces bayanus*, *Saccaromyces boulardi*, *Saccharomyces cerevisiae*, *Streptomyces fungicidicus*, *Streptomyces griseochromogenes*, *Streptomyces griseus*, *Streptomyces lividans*, *Streptomyces olivogriseus*, *Streptomyces rameus*, *Streptomyces tanashiensis*, *Streptomyces vinaceus*, *Trichoderma reesei* and *Xanthophyllomyces dendrorhous* (formerly *Phaffia rhodozyma*).

[0064] In general, if a eukaryotic cell is used, a non-pathogenic strain is preferred. Illustrative examples of non-pathogenic strains include but are not limited to: *Fusarium graminearum*, *Fusarium venenatum*, *Pichia pastoris*, *Saccharomyces boulardi*, and *Saccharomyces cerevisiae*.

[0065] In addition, certain strains have been designated by the Food and Drug Administration as GRAS or Generally Regarded As Safe. These strains include: *Bacillus subtilis*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, and *Saccharomyces cerevisiae*.

IPP Pathways

[0066] The host cells of the present invention comprise or utilize the MEV pathway, the DXP pathway or both to synthesize IPP and its isomer, DMAPP. In general, eukaryotes other than plants use the MEV isoprenoid pathway exclusively to convert acetyl-CoA to IPP, which is subsequently isomerized to DMAPP.

Prokaryotes, with some exceptions, use the mevalonate-independent or DXP pathway to produce IPP and DMAPP separately through a branch point. In general, plants use both the MEV and DXP pathways for IPP synthesis.

MEV Pathway

[0067] A schematic representation of the MEV pathway is described in Figure 1A. In general, the pathway comprises six steps.

[0068] In the first step, two molecules of acetyl-coenzyme A are enzymatically combined to form acetoacetyl-CoA. An enzyme known to catalyze this step is, for example, acetyl-CoA thiolase (also known as acetyl-CoA acetyltransferase). Illustrative examples of nucleotide sequences include but are not limited to the following GenBank accession numbers and the organism from which the sequences derived: (NC_000913 REGION: 2324131..2325315; *Escherichia coli*), (D49362; *Paracoccus denitrificans*), and (L20428; *Saccharomyces cerevisiae*).

[0069] In the second step of the MEV pathway, acetoacetyl-CoA is enzymatically condensed with another molecule of acetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). An enzyme known to catalyze this step is, for example, HMG-CoA synthase. Illustrative examples of nucleotide sequences include but are not limited to: (NC_001145: complement 19061..20536; *Saccharomyces cerevisiae*), (X96617; *Saccharomyces cerevisiae*), (X83882; *Arabidopsis thaliana*), (AB037907; *Kitasatospora griseola*), (BT007302; *Homo sapiens*), and (NC_002758, Locus tag SAV2546, GeneID 1122571; *Staphylococcus aureus*).

[0070] In the third step, HMG-CoA is enzymatically converted to mevalonate. An enzyme known to catalyze this step is, for example, HMG-CoA reductase. Illustrative examples of nucleotide sequences include but are not limited to: (NM_206548; *Drosophila melanogaster*), (NC_002758, Locus tag SAV2545, GeneID 1122570; *Staphylococcus aureus*), (NM_204485; *Gallus gallus*), (AB015627; *Streptomyces sp.* KO 3988), (AF542543; *Nicotiana attenuata*), (AB037907; *Kitasatospora griseola*), (AX128213, providing the sequence encoding a truncated HMGR; *Saccharomyces cerevisiae*), and (NC_001145: complement (115734..118898; *Saccharomyces cerevisiae*).

[0071] In the fourth step, mevalonate is enzymatically phosphorylated to form mevalonate 5-phosphate. An enzyme known to catalyze this step is, for example, mevalonate kinase. Illustrative examples of nucleotide sequences include but are not limited to: (L77688; *Arabidopsis thaliana*), and (X55875; *Saccharomyces cerevisiae*).

[0072] In the fifth step, a second phosphate group is enzymatically added to mevalonate 5-phosphate to form mevalonate 5-pyrophosphate. An enzyme known to catalyze this step is, for example, phosphomevalonate

kinase. Illustrative examples of nucleotide sequences include but are not limited to: (AF429385; *Hevea brasiliensis*), (NM_006556; *Homo sapiens*), and (NC_001145, complement 712315..713670; *Saccharomyces cerevisiae*).

- [0073] In the sixth step, mevalonate 5-pyrophosphate is enzymatically converted into IPP. An enzyme known to catalyze this step is, for example, mevalonate pyrophosphate decarboxylase. Illustrative examples of nucleotide sequences include but are not limited to: (X97557; *Saccharomyces cerevisiae*), (AF290095; *Enterococcus faecium*), and (U49260; *Homo sapiens*).
- [0074] If IPP is to be converted to DMAPP, then a seventh step is required. An enzyme known to catalyze this step is, for example, IPP isomerase. Illustrative examples of nucleotide sequences include but are not limited to: (NC_000913, 3031087..3031635; *Escherichia coli*), and (AF082326; *Haematococcus pluvialis*). If the conversion to DMAPP is required, an increased expression of IPP isomerase ensures that the conversion of IPP into DMAPP does not represent a rate-limiting step in the overall pathway.

DXP Pathway

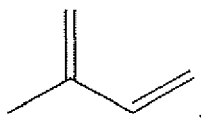
- [0075] A schematic representation of the DXP pathway is described in Figure 1B. In general, the DXP pathway comprises seven steps. In the first step, pyruvate is condensed with D-glyceraldehyde 3-phosphate to make 1-deoxy-D-xylulose-5-phosphate. An enzyme known to catalyze this step is, for example, 1-deoxy-D-xylulose-5-phosphate synthase. Illustrative examples of nucleotide sequences include but are not limited to: (AF035440; *Escherichia coli*), (NC_002947, locus tag PP0527; *Pseudomonas putida* KT2440), (CP000026, locus tag SPA2301; *Salmonella enterica* Paratyphi, see ATCC 9150), (NC_007493, locus tag RSP_0254; *Rhodobacter sphaeroides* 2.4.1), (NC_005296, locus tag RPA0952; *Rhodospseudomonas palustris* CGA009), (NC_004556, locus tag PD1293; *Xylella fastidiosa* Temecula1), and (NC_003076, locus tag AT5G11380; *Arabidopsis thaliana*).
- [0076] In the second step, 1-deoxy-D-xylulose-5-phosphate is converted to 2C-methyl-D-erythritol-4-phosphate. An enzyme known to catalyze this step is, for example, 1-deoxy-D-xylulose-5-phosphate reductoisomerase. Illustrative examples of nucleotide sequences include but are not limited to: (AB013300; *Escherichia coli*), (AF148852; *Arabidopsis thaliana*), (NC_002947, locus tag PP1597; *Pseudomonas putida* KT2440), (AL939124, locus tag SCO5694; *Streptomyces coelicolor* A3(2)), (NC_007493, locus tag RSP_2709; *Rhodobacter sphaeroides* 2.4.1), and (NC_007492, locus tag Pfl_1107; *Pseudomonas fluorescens* PfO-1).
- [0077] In the third step, 2C-methyl-D-erythritol-4-phosphate is converted to 4-diphosphocytidyl-2C-methyl-D-erythritol. An enzyme known to catalyze this step is, for example, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase. Illustrative examples of nucleotide sequences include but are not limited to: (AF230736; *Escherichia coli*), (NC_007493, locus_tag RSP_2835; *Rhodobacter sphaeroides* 2.4.1), (NC_003071, locus_tag AT2G02500; *Arabidopsis thaliana*), and (NC_002947, locus_tag PP1614; *Pseudomonas putida* KT2440).
- [0078] In the fourth step, 4-diphosphocytidyl-2C-methyl-D-erythritol is converted to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate. An enzyme known to catalyze this step is, for example, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase. Illustrative examples of nucleotide sequences include but are not limited to: (AF216300; *Escherichia coli*) and (NC_007493, locus_tag RSP_1779; *Rhodobacter sphaeroides* 2.4.1).
- [0079] In the fifth step, 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate is converted to 2C-methyl-D-erythritol 2, 4-cyclodiphosphate. An enzyme known to catalyze this step is, for example, 2C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase. Illustrative examples of nucleotide sequences include but are not

limited to: (AF230738; *Escherichia coli*), (NC_007493, locus_tag RSP_6071; *Rhodobacter sphaeroides* 2.4.1), and (NC_002947, locus_tag PP1618; *Pseudomonas putida* KT2440).

- [0080] In the sixth step, 2C-methyl-D-erythritol 2, 4-cyclodiphosphate is converted to 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate. An enzyme known to catalyze this step is, for example, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase. Illustrative examples of nucleotide sequences include but are not limited to: (AY033515; *Escherichia coli*), (NC_002947, locus_tag PP0853; *Pseudomonas putida* KT2440), and (NC_007493, locus_tag RSP_2982; *Rhodobacter sphaeroides* 2.4.1).
- [0081] In the seventh step, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate is converted into either IPP or its isomer, DMAPP. An enzyme known to catalyze this step is, for example, isopentyl/dimethylallyl diphosphate synthase. Illustrative examples of nucleotide sequences include but are not limited to: (AY062212; *Escherichia coli*) and (NC_002947, locus_tag PP0606; *Pseudomonas putida* KT2440).
- [0082] In some embodiments, "cross talk" (or interference) between the host cell's own metabolic processes and those processes involved with the production of IPP as provided by the present invention are minimized or eliminated entirely. For example, cross talk is minimized or eliminated entirely when the host microorganism relies exclusively on the DXP pathway for synthesizing IPP, and a MEV pathway is introduced to provide additional IPP. Such a host organisms would not be equipped to alter the expression of the MEV pathway enzymes or process the intermediates associated with the MEV pathway. Organisms that rely exclusively or predominately on the DXP pathway include, for example, *Escherichia coli*.
- [0083] In some embodiments, the host cell produces IPP via the MEV pathway, either exclusively or in combination with the DXP pathway. In other embodiments, a host's DXP pathway is functionally disabled so that the host cell produces IPP exclusively through a heterologously introduced MEV pathway. The DXP pathway can be functionally disabled by disabling gene expression or inactivating the function of one or more of the DXP pathway enzymes.

C₅ Compounds

- [0084] *C₅* compounds of the invention generally are derived from IPP or DMAPP. These compounds are also known as hemiterpenes because they are derived from a single isoprene unit (IPP or DMAPP).
- [0085] Isoprene
- [0086] Isoprene, whose structure is



is found in many plants. Isoprene is made from IPP by isoprene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (AB198190; *Populus alba*) and (AJ294819; *Populus alba* x *Populus tremula*).

C₁₀ Compounds

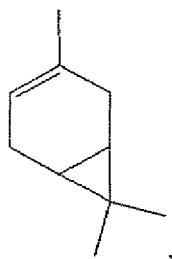
- [0087] *C₁₀* compounds of the invention generally derived from geranyl pyrophosphate (GPP) which is made by the condensation of IPP with DMAPP. An enzyme known to catalyze this step is, for example, geranyl pyrophosphate synthase. These *C₁₀* compounds are also known as monoterpenes because they are derived from two isoprene units.
- [0088] Figure 2 shows schematically how IPP and DMAPP can produce GPP, which can be further processed to a monoterpene.

[0089] Illustrative examples of nucleotide sequences for geranyl pyrophosphate synthase include but are not limited to: (AF513111; *Abies grandis*), (AF513112; *Abies grandis*), (AF513113; *Abies grandis*), (AY534686; *Antirrhinum majus*), (AY534687; *Antirrhinum majus*), (Y17376; *Arabidopsis thaliana*), (AE016877, Locus AP11092; *Bacillus cereus*; ATCC 14579), (AJ243739; *Citrus sinensis*), (AY534745; *Clarkia breweri*), (AY953508; *Ips pini*), (DQ286930; *Lycopersicon esculentum*), (AF182828; *Mentha x piperita*), (AF182827; *Mentha x piperita*), (MPI249453; *Mentha x piperita*), (PZE431697, Locus CAD24425; *Paracoccus zeaxanthinifaciens*), (AY866498; *Picrorhiza kurrooa*), (AY351862; *Vitis vinifera*), and (AF203881, Locus AAF12843; *Zymomonas mobilis*).

[0090] GPP is then subsequently converted to a variety of C₁₀ compounds. Illustrative examples of C₁₀ compounds include but are not limited:

[0091] Carene

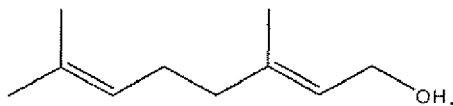
[0092] Carene, whose structure is



is found in the resin of many trees, particularly pine trees. Carene is made from GPP from carene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (AF461460, REGION 43..1926; *Picea abies*) and (AF527416, REGION: 78..1871; *Salvia stenophylla*).

[0093] Geraniol

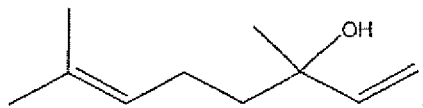
[0094] Geraniol (also known as rhodnol), whose structure is



is the main component of oil-of-rose and palmarosa oil. It also occurs in geranium, lemon, and citronella. Geraniol is made from GPP by geraniol synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (AJ457070; *Cinnamomum tenuipilum*), (AY362553; *Ocimum basilicum*), (DQ234300; *Perilla frutescens* strain 1864), (DQ234299; *Perilla citriodora* strain 1861), (DQ234298; *Perilla citriodora* strain 4935), and (DQ088667; *Perilla citriodora*)

[0095] Linalool

[0096] Linalool, whose structure is

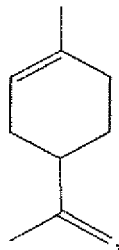


is found in many flowers and spice plants such as coriander seeds. Linalool is made from GPP by linalool synthase. Illustrative examples of a suitable nucleotide sequence include but are not limited to: (AF497485; *Arabidopsis thaliana*), (AC002294, Locus AAB71482; *Arabidopsis thaliana*), (AY059757; *Arabidopsis thaliana*), (NM_104793; *Arabidopsis thaliana*), (AF154124; *Artemisia annua*), (AF067603; *Clarkia breweri*), (AF067602; *Clarkia concinna*), (AF067601; *Clarkia breweri*), (U58314; *Clarkia breweri*), (AY840091; *Lycopersicon esculentum*), (DQ263741; *Lavandula angustifolia*), (AY083653;

Mentha citrate), (AY693647; *Ocimum basilicum*), (XM_463918; *Oryza sativa*), (AP004078, Locus BAD07605; *Oryza sativa*), (XM_463918, Locus XP_463918; *Oryza sativa*), (AY917193; *Perilla citriodora*), (AF271259; *Perilla frutescens*), (AY473623; *Picea abies*), (DQ195274; *Picea sitchensis*), and (AF444798; *Perilla frutescens* var. *crispa* cultivar No. 79).

[0097] Limonene

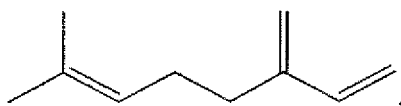
[0098] Limonene, whose structure is



is found in the rind of citrus fruits and peppermint. Limonene is made from GPP by limonene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (+)-limonene synthases (AF514287, REGION: 47..1867; *Citrus limon*) and (AY055214, REGION: 48..1889; *Agastache rugosa*) and (-)-limonene synthases (DQ195275, REGION: 1..1905; *Picea sitchensis*), (AF006193, REGION: 73..1986; *Abies grandis*), and (MHC4SLSP, REGION: 29..1828; *Mentha spicata*).

[0099] Myrcene

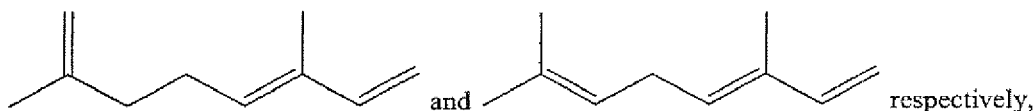
[00100] Myrcene, whose structure is



is found in the essential oil in many plants including bay, verbena, and myrcia from which it gets its name. Myrcene is made from GPP by myrcene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (U87908; *Abies grandis*), (AY195609; *Antirrhinum majus*), (AY195608; *Antirrhinum majus*), (NM_127982; *Arabidopsis thaliana* TPS10), (NM_113485; *Arabidopsis thaliana* ATTPS-CIN), (NM_113483; *Arabidopsis thaliana* ATTPS-CIN), (AF271259; *Perilla frutescens*), (AY473626; *Picea abies*), (AF369919; *Picea abies*), and (AJ304839; *Quercus ilex*).

[00101] Ocimene

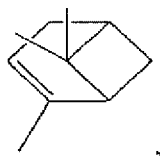
[00102] α - and β -Ocimene, whose structures are



are found in a variety of plants and fruits including *Ocimum basilicum* and is made from GPP by ocimene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (AY195607; *Antirrhinum majus*), (AY195609; *Antirrhinum majus*), (AY195608; *Antirrhinum majus*), (AK221024; *Arabidopsis thaliana*), (NM_113485; *Arabidopsis thaliana* ATTPS-CIN), (NM_113483; *Arabidopsis thaliana* ATTPS-CIN), (NM_117775; *Arabidopsis thaliana* ATTPS03), (NM_001036574; *Arabidopsis thaliana* ATTPS03), (NM_127982; *Arabidopsis thaliana* TPS10), (AB110642; *Citrus unshiu* CitMTSL4), and (AY575970; *Lotus corniculatus* var. *japonicus*).

[00103] α -Pinene

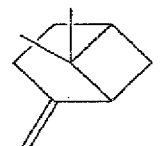
[00104] α -Pinene, whose structure is



is found in pine trees and eucalyptus. α -Pinene is made from GPP by α -pinene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (+) α -pinene synthase (AF543530, REGION: 1..1887; *Pinus taeda*), (-) α -pinene synthase (AF543527, REGION: 32..1921; *Pinus taeda*), and (+)/(-) α -pinene synthase (AGU87909, REGION: 6111892; *Abies grandis*).

[00105] β -Pinene

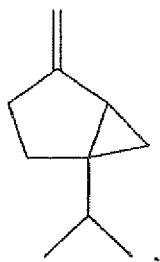
[00106] β -Pinene, whose structure is



is found in pine trees, rosemary, parsley, dill, basil, and rose. β -Pinene is made from GPP by β -pinene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to: (-) β -pinene synthases (AF276072, REGION: 1..1749; *Artemisia annua*) and (AF514288, REGION: 26..1834; *Citrus limon*).

[00107] Sabinene

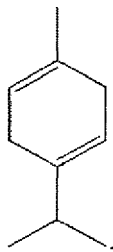
[00108] Sabinene, whose structure is



is found in black pepper, carrot seed, sage, and tea trees. Sabinene is made from GPP by sabinene synthase. An illustrative example of a suitable nucleotide sequence includes but is not limited to AF051901, REGION: 26..1798 from *Salvia officinalis*.

[00109] γ -Terpinene

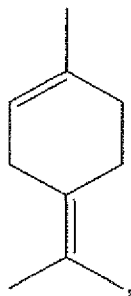
[00110] γ -Terpinene, whose structure is



is a constituent of the essential oil from citrus fruits. Biochemically, γ -terpinene is made from GPP by a γ -terpinene synthase. Illustrative examples of suitable nucleotide sequences include: (AF514286, REGION: 30..1832 from *Citrus limon*) and (AB110640, REGION 1..1803 from *Citrus unshiu*).

[00111] Terpinolene

[00112] Terpinolene, whose structure is



is found in black currant, cypress, guava, lychee, papaya, pine, and tea. Terpinolene is made from GPP by terpinolene synthase. An illustrative example of a suitable nucleotide sequence includes but is not limited to AY906866, REGION: 10..1887 from *Pseudotsuga menziesii*.

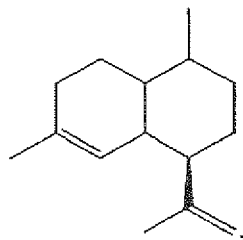
C₁₅ Compounds

- [00113] C_{15} compounds of the invention generally derive from farnesyl pyrophosphate (FPP) which is made by the condensation of two molecules of IPP with one molecule of DMAPP. An enzyme known to catalyze this step is, for example, farnesyl pyrophosphate synthase. These C_{15} compounds are also known as sesquiterpenes because they are derived from three isoprene units.
- [00114] Figure 2 shows schematically how IPP and DMAPP can be combined to produce FPP, which can be further processed to a sesquiterpene.
- [00115] Illustrative examples of nucleotide sequences for farnesyl pyrophosphate synthase include but are not limited to: (ATU80605; *Arabidopsis thaliana*), (ATHFPS2R; *Arabidopsis thaliana*), (AAU36376; *Artemisia annua*), (AF461050; *Bos taurus*), (D00694; *Escherichia coli* K-12), (AE009951, Locus AAL95523; *Fusobacterium nucleatum subsp. nucleatum* ATCC 25586), (GFFPPSGEN; *Gibberella fujikuroi*), (CP000009, Locus AAW60034; *Gluconobacter oxydans* 621H), (AF019892; *Helianthus annuus*), (HUMFAPS; *Homo sapiens*), (KLPFPSQCR; *Kluyveromyces lactis*), (LAU15777; *Lupinus albus*), (LAU20771; *Lupinus albus*), (AF309508; *Mus musculus*), (NCFPPSGEN; *Neurospora crassa*), (PAFPS1; *Parthenium argentatum*), (PAFPS2; *Parthenium argentatum*), (RATFAPS; *Rattus norvegicus*), (YSCFPP; *Saccharomyces cerevisiae*), (D89104; *Schizosaccharomyces pombe*), (CP000003, Locus AAT87386; *Streptococcus pyogenes*), (CP000017, Locus AAZ51849; *Streptococcus pyogenes*), (NC_008022, Locus YP_598856; *Streptococcus pyogenes* MGAS10270), (NC_008023, Locus YP_600845; *Streptococcus pyogenes* MGAS2096), (NC_008024, Locus YP_602832; *Streptococcus pyogenes* MGAS10750), and (MZEFP; *Zea mays*).
- [00116] Alternatively, FPP can also be made by adding IPP to GPP. Illustrative examples of nucleotide sequences encoding for an enzyme capable of this reaction include but are not limited to: (AE000657, Locus AAC06913; *Aquifex aeolicus* VF5), (NM_202836; *Arabidopsis thaliana*), (D84432, Locus BAA12575; *Bacillus subtilis*), (U12678, Locus AAC28894; *Bradyrhizobium japonicum* USDA 110), (BACFDPS; *Geobacillus stearothermophilus*), (NC_002940, Locus NP_873754; *Haemophilus ducreyi* 35000HP), (L42023, Locus AAC23087; *Haemophilus influenzae* Rd KW20), (J05262; *Homo sapiens*), (YP_395294; *Lactobacillus sakei subsp. sakei* 23K), (NC_005823, Locus YP_000273; *Leptospira interrogans* serovar Copenhageni str. Fiocruz L1-130), (AB003187; *Micrococcus luteus*), (NC_002946, Locus YP_208768; *Neisseria gonorrhoeae* FA 1090), (U00090, Locus AAB91752; *Rhizobium sp.* NGR234), (J05091; *Saccharomyces cerevisiae*), (CP000031, Locus AAV93568; *Silicibacter pomeroyi* DSS-3), (AE008481, Locus AAK99890; *Streptococcus pneumoniae* R6), and (NC_004556, Locus NP_779706; *Xylella fastidiosa* Temecula1).

[00117] FPP is then subsequently converted to a variety of C₁₅ compounds. Illustrative examples of C₁₅ compounds include but are not limited to:

[00118] Amorphadiene

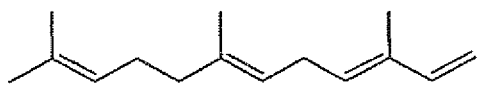
[00119] Amorphadiene, whose structure is



is a precursor to artemisinin which is made by *Artemisia anna*. Amorphadiene is made from FPP by amorphadiene synthase. An illustrative example of a suitable nucleotide sequence is SEQ ID NO. 37 of U.S. Patent No. 7,192,751.

[00120] α -Farnesene

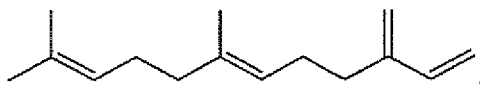
[00121] α -Farnesene, whose structure is



is found in various biological sources including but not limited to the Dufour's gland in ants and in the coating of apple and pear peels. α -Farnesene is made from FPP by α -farnesene synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to DQ309034 from *Pyrus communis* cultivar *d'Anjou* (pear; gene name AFS1) and AY182241 from *Malus domestica* (apple; gene AFS1). Pechous et al., *Planta* **219**(1):84-94 (2004).

[00122] β -Farnesene

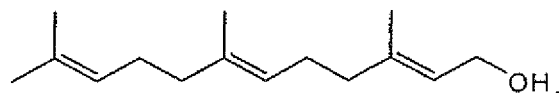
[00123] β -Farnesene, whose structure is



is found in various biological sources including but not limited to aphids and essential oils such as from peppermint. In some plants such as wild potato, β -farnesene is synthesized as a natural insect repellent. β -Farnesene is made from FPP by β -farnesene synthase. Illustrative examples of suitable nucleotide sequences include but is not limited to GenBank accession number AF024615 from *Mentha x piperita* (peppermint; gene Tspa11), and AY835398 from *Artemisia annua*. Picaud et al., *Phytochemistry* **66**(9): 961-967 (2005).

[00124] Farnesol

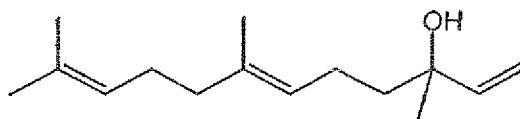
[00125] Farnesol, whose structure is



is found in various biological sources including insects and essential oils such as from citronella, neroli, cyclamen, lemon grass, tuberose, and rose. Farnesol is made from FPP by a hydroxylase such as farnesol synthase. Illustrative examples of suitable nucleotide sequences include but are not limited to GenBank accession number AF529266 from *Zea mays* and YDR481C from *Saccharomyces cerevisiae* (gene Pho8). Song, L., *Applied Biochemistry and Biotechnology* **128**:149-158 (2006).

[00126] Nerolidol

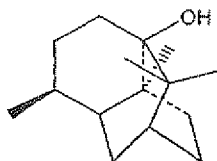
[00127] Nerolidol, whose structure is



is also known as peruvicol, and is found in various biological sources including as essential oils such as from neroli, ginger, jasmine, lavender, tea tree, and lemon grass. Nerolidol is made from FPP by a hydroxylase such as nerolidol synthase. An illustrative example of a suitable nucleotide sequence includes but is not limited to AF529266 from *Zea mays* (maize; gene tps1).

[00128] Patchouliol

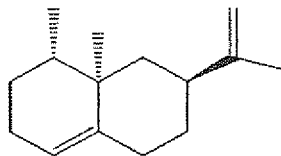
[00129] Patchouliol, whose structure is



is also known as patchouli alcohol and is a constituent of the essential oil of *Pogostemon patchouli*. Patchouliol is made from FPP by patchouliol synthase. An illustrative example of a suitable nucleotide sequence includes but is not limited to AY508730 REGION: 1..1659 from *Pogostemon cablin*.

[00130] Valencene

[00131] Valencene, whose structure is



is one of the main chemical components of the smell and flavour of oranges and is found in orange peels. Valencene is made from FPP by nootkatone synthase. Illustrative examples of a suitable nucleotide sequence includes but is not limited to AF441124 REGION: 1..1647 from *Citrus sinensis* and AY917195 REGION: 1..1653 from *Perilla frutescens*.

***C*₂₀ Compounds**

[00132] *C*₂₀ compounds of the invention generally derived from geranylgeraniol pyrophosphate (GGPP) which is made by the condensation of three molecules of IPP with one molecule of DMAPP. An enzyme known to catalyze this step is, for example, geranylgeranyl pyrophosphate synthase. These *C*₂₀ compounds are also known as diterpenes because they are derived from four isoprene units.

[00133] Figure 2 shows schematically how IPP and DMAPP can be combined to produce GGPP, which can be further processed to a diterpene, or can be further processed to produce a carotenoid.

[00134] Illustrative examples of nucleotide sequences for geranylgeranyl pyrophosphate synthase include but are not limited to: (ATHGERPYRS; *Arabidopsis thaliana*), (BT005328; *Arabidopsis thaliana*), (NM_119845; *Arabidopsis thaliana*), (NZ_AAJM01000380, Locus ZP_00743052; *Bacillus thuringiensis* serovar *israelensis*, ATCC 35646 sq1563), (CRGGPPS; *Catharanthus roseus*), (NZ_AABF02000074, Locus ZP_00144509; *Fusobacterium nucleatum* subsp. *vincentii*, ATCC 49256), (GFGGPPSGN; *Gibberella*

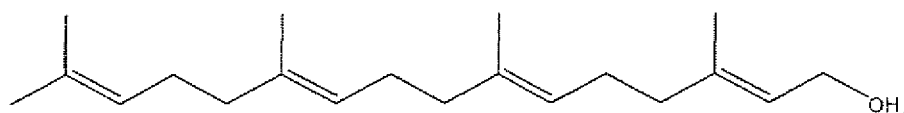
fujikuroi), (AY571321; *Ginkgo biloba*), (AB055496; *Hevea brasiliensis*), (AB017971; *Homo sapiens*), (MCI276129; *Mucor circinelloides f. lusitanicus*), (AB016044; *Mus musculus*), (AABX01000298, Locus NCU01427; *Neurospora crassa*), (NCU20940; *Neurospora crassa*), (NZ_AAKL01000008, Locus ZP_00943566; *Ralstonia solanacearum* UW551), (AB118238; *Rattus norvegicus*), (SCU31632; *Saccharomyces cerevisiae*), (AB016095; *Synechococcus elongates*), (SAGGPS; *Sinapis alba*), (SSOGDS; *Sulfolobus acidocaldarius*), (NC_007759, Locus YP_461832; *Syntrophus aciditrophicus* SB), and (NC_006840, Locus YP_204095; *Vibrio fischeri* ES114).

[00135] Alternatively, GGPP can also be made by adding IPP to FPP. Illustrative examples of nucleotide sequences encoding an enzyme capable of this reaction include but are not limited to: (NM_112315; *Arabidopsis thaliana*), (ERWCRT; *Pantoea agglomerans*), (D90087, Locus BAA14124; *Pantoea ananatis*), (X52291, Locus CAA36538; *Rhodobacter capsulatus*), (AF195122, Locus AAF24294; *Rhodobacter sphaeroides*), and (NC_004350, Locus NP_721015; *Streptococcus mutans* UA159).

[00136] GGPP is then subsequently converted to a variety of C₂₀ isoprenoids. Illustrative examples of C₂₀ compounds include but are not limited to:

[00137] Geranylgeraniol

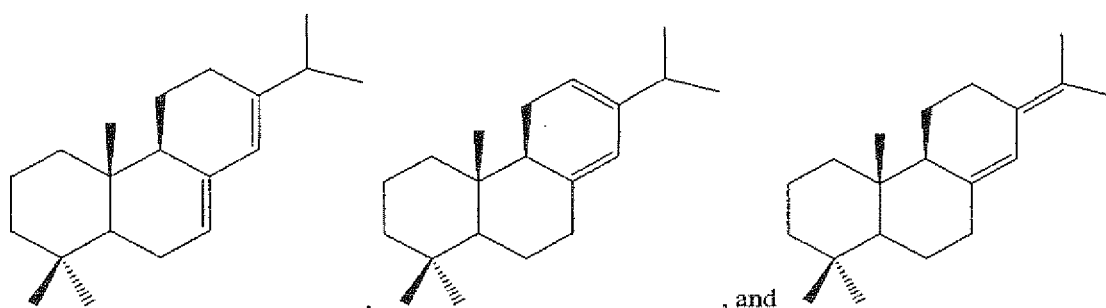
[00138] Geranylgeraniol, whose structure is



is a constituent of wood oil from *Cedrela toona* and of linseed oil. Geranylgeraniol can be made by e.g., adding to the expression constructs a phosphatase gene after the gene for a GGPP synthase.

[00139] Abietadiene

[00140] Abietadiene encompasses the following isomers:



and is found in trees such as *Abies grandis*. Abietadiene is made by abietadiene synthase. An illustrative example of a suitable nucleotide sequence includes but are not limited to: (U50768; *Abies grandis*) and (AY473621; *Picea abies*).

C₂₀₊ Compounds

[00141] C₂₀₊ compounds are also within the scope of the present invention. Illustrative examples of such compounds include sesterterpenes (C₂₅ compound made from five isoprene units), triterpenes (C₃₀ compounds made from six isoprene units), and tetraterpenes (C₄₀ compound made from eight isoprene units). These compounds are made by using similar methods described herein and substituting or adding nucleotide sequences for the appropriate synthase(s).

Engineering Pathways

[00142] The present invention utilizes an engineered MEV and/or DXP pathway to effect the high-level production of isoprenoids in a host cell. The pathway is typically engineered via recombinant DNA technology by expressing heterologous sequences encoding enzymes in at least one of these pathways.

- [00143] The subject nucleotide acids can be expressed by a single or multiple vectors. For example, a single expression vector can comprise at least two, three, four, five, or all of the heterologous sequences encoding the entire MEV or DXP pathway enzymes. While the choice of single or multiple vectors may depend on the size of the heterologous sequences and the capacity of the vectors, it will largely dependent on the overall yield of a given isoprenoid that the vector is able to provide when expressed in a selected host cell. The subject vectors can stay replicable episomally, or as an integral part of the host cell genome. Typically, the latter is preferred for a sustained propagation of the host cell.
- [00144] In certain host cells, the one or more heterologous sequences encoding the MEV or DXP pathway enzymes may be controlled by one or more operons. In some instances, a two or three operon system provides a higher yield of an isoprenoid over a single operon system.
- [00145] Where desired, the subject nucleic acid sequences can be modified to reflect the codon preference of a selected host cell to effect a higher expression of such sequences in a host cell. For example, the subject nucleotide sequences will in some embodiments be modified for yeast codon preference. See, e.g., Bennetzen and Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. As another non-limiting example, the nucleotide sequences will in other embodiments be modified for *E. coli* codon preference. See, e.g., Gouy and Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. See also Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292. Codon usage tables for many organisms are available, which can be used as a reference in designing sequences of the present invention. The use of prevalent codons of a given host microorganism generally increases the likelihood of translation, and hence the expression level of the desired sequences.
- [00146] Preparation of the subject nucleic acids can be carried out by a variety of routine recombinant techniques and synthetic procedures. Briefly, the subject nucleic acids can be prepared genomic DNA fragments, cDNAs, and RNAs, all of which can be extracted directly from a cell or recombinantly produced by various amplification processes including but not limited to PCR and rt-PCR.
- [00147] Direct chemical synthesis of nucleic acids typically involves sequential addition of 3'-blocked and 5'-blocked nucleotide monomers to the terminal 5'-hydroxyl group of a growing nucleotide polymer chain, wherein each addition is effected by nucleophilic attack of the terminal 5'-hydroxyl group of the growing chain on the 3'-position of the added monomer, which is typically a phosphorus derivative, such as a phosphotriester, phosphoramidite, or the like. Such methodology is known to those of ordinary skill in the art and is described in the pertinent texts and literature (for example, Matteucci et al. (1980) *Tet. Lett.* 521:719; U.S. Pat. No. 4,500,707 to Caruthers et al.; and U.S. Pat. Nos. 5,436,327 and 5,700,637 to Southern et al.).
- [00148] The level of transcription of a nucleic acid in a host microorganism can be increased in a number of ways. For example, this can be achieved by increasing the copy number of the nucleotide sequence encoding the enzyme (e.g., by using a higher copy number expression vector comprising a nucleotide sequence encoding the enzyme, or by introducing additional copies of a nucleotide sequence encoding the enzyme into the genome of the host microorganism, for example, by recA-mediated recombination, use of "suicide" vectors, recombination using lambda phage recombinase, and/or insertion via a transposon or transposable element). In addition, it can be carried out by changing the order of the coding regions on the polycistronic mRNA of an operon or breaking up an operon into individual genes, each with its own control elements, or increasing the strength of the promoter (transcription initiation or transcription control sequence) to which the enzyme coding region is operably linked (for example, using a consensus arabinose- or lactose-

inducible promoter in an *Escherichia coli* host microorganism in place of a modified lactose-inducible promoter, such as the one found in pBluescript and the pBBR1MCS plasmids), or using an inducible promoter and inducing the inducible-promoter by adding a chemical to a growth medium. The level of translation of a nucleotide sequence in a host microorganism can be increased in a number of ways, including, but not limited to, increasing the stability of the mRNA, modifying the sequence of the ribosome binding site, modifying the distance or sequence between the ribosome binding site and the start codon of the enzyme coding sequence, modifying the entire intercistronic region located "upstream of" or adjacent to the 5' side of the start codon of the enzyme coding region, stabilizing the 3'-end of the mRNA transcript using hairpins and specialized sequences, modifying the codon usage of enzyme, altering expression of rare codon tRNAs used in the biosynthesis of the enzyme, and/or increasing the stability of the enzyme, as, for example, via mutation of its coding sequence. Determination of preferred codons and rare codon tRNAs can be based on a sequence analysis of genes derived from the host microorganism.

[00149] The activity of a MEV, DXP, or prenyltransferase in a host can be altered in a number of ways, including, but not limited to, expressing a modified form of the enzyme that exhibits increased solubility in the host cell, expressing an altered form of the enzyme that lacks a domain through which the activity of the enzyme is inhibited, expressing a modified form of the enzyme that has a higher Kcat or a lower Km for the substrate, or expressing an altered form of the enzyme that is not affected by feed-back or feed-forward regulation by another molecule in the pathway. Such variant enzymes can also be isolated through random mutagenesis of a broader specificity enzyme, as described below, and a nucleotide sequence encoding such variant enzyme can be expressed from an expression vector or from a recombinant gene integrated into the genome of a host microorganism.

[00150] The subject vector can be constructed to yield a desired level of copy numbers of the encoded enzyme. In some embodiments, the subject vectors yield at least 10, between 10 to 20, between 20-50, between 50-100, or even higher than 100 copies of the HMG-CoA reductase, mevalonate kinase, or both. Low copy number plasmids generally provide fewer than about 20 plasmid copies per cell; medium copy number plasmids generally provide from about 20 plasmid copies per cell to about 50 plasmid copies per cell, or from about 20 plasmid copies per cell to about 80 plasmid copies per cell; and high copy number plasmids generally provide from about 80 plasmid copies per cell to about 200 plasmid copies per cell, or more.

[00151] Suitable low copy expression vectors for *Escherichia coli* include, but are not limited to, pACYC184, pBeloBac11, pBR332, pBAD33, pBBR1MCS and its derivatives, pSC101, SuperCos (cosmid), and pWE15 (cosmid). Suitable medium copy expression vectors for *Escherichia coli* include, but are not limited to pTrc99A, pBAD24, and vectors containing a ColE1 origin of replication and its derivatives. Suitable high copy number expression vectors for *Escherichia coli* include, but are not limited to, pUC, pBluescript, pGEM, and pTZ vectors. Suitable low-copy (centromeric) expression vectors for yeast include, but are not limited to, pRS415 and pRS416 (Sikorski & Hieter (1989) Genetics 122:19-27). Suitable high-copy 2 micron expression vectors in yeast include, but are not limited to, pRS425 and pRS426 (Christainson et al. (1992) Gene 110:119-122). Alternative 2 micron expression vectors include non-selectable variants of the 2 micron vector (Bruschi & Ludwig (1988) Curr. Genet. 15:83-90) or intact 2 micron plasmids bearing an expression cassette (as exemplified in U.S. Pat. Appl. 20050084972) or 2 micron plasmids bearing a defective selection marker such as LEU2d (Erhanrt et al. (1983) J. Bacteriol. 156 (2): 625-635) or URA3d (Okkels (1996) Annals of the New York Academy of Sciences 782(1): 202-207).

- [00152] Regulatory elements include, for example, promoters and operators can also be engineered to increase the metabolic flux of the MEV or DXP pathways by increasing the expression of one or more genes that play a significant role in determining the overall yield of an isoprenoid produced. A promoter is a sequence of nucleotides that initiates and controls the transcription of a nucleic acid sequence by an RNA polymerase enzyme. An operator is a sequence of nucleotides adjacent to the promoter that functions to control transcription of the desired nucleic acid sequence. The operator contains a protein-binding domain where a specific repressor protein can bind. In the absence of a suitable repressor protein, transcription initiates through the promoter. In the presence of a suitable repressor protein, the repressor protein binds to the operator and thereby inhibits transcription from the promoter. Promoters and operators are also referred to as transcriptional regulators.
- [00153] In some embodiments of the present invention, promoters used in expression vectors are inducible. In other embodiments, the promoters used in expression vectors are constitutive. In some embodiments, one or more nucleic acid sequences are operably linked to an inducible promoter, and one or more other nucleic acid sequences are operably linked to a constitutive promoter.
- [00154] Non-limiting examples of suitable promoters for use in prokaryotic host cells include a bacteriophage T7 RNA polymerase promoter; a trp promoter; a lac operon promoter; a hybrid promoter, for example, a lac/tac hybrid promoter, a tac/trc hybrid promoter, a trp/lac promoter, a T7/lac promoter; a trc promoter; a tac promoter, and the like; an araBAD promoter; *in vivo* regulated promoters, such as an *ssaG* promoter or a related promoter (see, for example, U.S. Patent Publication No. 20040131637), a *pagC* promoter (Pulkkinen and Miller, J. Bacteriol. (1991) 173(1):86-93; Alpuche-Aranda et al. (1992) Proc. Natl. Acad. Sci. U S A. 89(21):10079-83), a *nirB* promoter (Harborne et al. (1992) Mol. Micro. 6:2805-2813), and the like (see, for example, Dunstan et al. (1999) Infect. Immun. 67:5133-5141; McKelvie et al. (2004) Vaccine 22:3243-3255; and Chatfield et al. (1992) Biotechnol. 10:888-892); a sigma70 promoter, for example, a consensus sigma70 promoter (see, for example, GenBank Accession Nos. AX798980, AX798961, and AX798183); a stationary phase promoter, for example, a *dps* promoter, an *spv* promoter, and the like; a promoter derived from the pathogenicity island SPI-2 (see, for example, WO96/17951); an actA promoter (see, for example, Shetron-Rama et al. (2002) Infect. Immun. 70:1087-1096); an rpsM promoter (see, for example, Valdivia and Falkow (1996) Mol. Microbiol. 22:367-378); a tet promoter (see, for example, Hillen et al. (1989) In Saenger W. and Heinemann U. (eds) Topics in Molecular and Structural Biology, Protein-Nucleic Acid Interaction. Macmillan, London, UK, Vol. 10, pp. 143-162); an SP6 promoter (see, for example, Melton et al. (1984) Nucl. Acids Res. 12:7035-7056); and the like.
- [00155] In some embodiment, the total activity of a heterologous MEV or DXP enzyme that plays a larger role in the overall yield of an isoprenoid relative to other enzymes in the respective pathways is increased by expressing the enzyme from a strong promoter. Suitable strong promoters for *Escherichia coli* include, but are not limited to Trc, Tac, T5, T7, and P_{Lambda}. In another embodiment of the present invention, the total activity of the one or more MEV pathway enzymes in a host is increased by expressing the enzyme from a strong promoter on a high copy number plasmid. Suitable examples, for *Escherichia coli* include, but are not limited to using Trc, Tac, T5, T7, and P_{Lambda} promoters with pBAD24, pBAD18, pGEM, pBluescript, pUC, and pTZ vectors.
- [00156] Non-limiting examples of suitable promoters for use in eukaryotic host cells include, but are not limited to, a CMV immediate early promoter, an HSV thymidine kinase promoter, an early or late SV40 promoter, LTRs from retroviruses, and a mouse metallothionein-I promoter.

- [00157] Non-limiting examples of suitable constitutive promoters for use in prokaryotic host cells include a sigma70 promoter (for example, a consensus sigma70 promoter). Non-limiting examples of suitable inducible promoters for use in bacterial host cells include the pL of bacteriophage λ ; Plac; P_{trp}; P_{tac} (P_{trp}-lac hybrid promoter); an isopropyl-beta-D44 thiogalactopyranoside (IPTG)-inducible promoter, for example, a *lacZ* promoter; a tetracycline inducible promoter; an arabinose inducible promoter, for example, PBAD (see, for example, Guzman et al. (1995) *J. Bacteriol.* 177:4121-4130); a xylose-inducible promoter, for example, P_{xyl} (see, for example, Kim et al. (1996) *Gene* 181:71-76); a GAL1 promoter; a tryptophan promoter; a lac promoter; an alcohol-inducible promoter, for example, a methanol-inducible promoter, an ethanol-inducible promoter; a raffinose-inducible promoter; a heat-inducible promoter, for example, heat inducible lambda PL promoter; a promoter controlled by a heat-sensitive repressor (for example, CI857-repressed lambda-based expression vectors; see, for example, Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177(2):327-34); and the like.
- [00158] Non-limiting examples of suitable constitutive promoters for use in yeast host cells include an ADH1, an ADH2, a PGK, or a LEU2 promoter. Non-limiting examples of suitable inducible promoters for use in yeast host cells include, but are not limited to, a divergent galactose-inducible promoter such as a GAL 1 or a GAL 10 promoter (West et al. (1984) *Mol. Cell. Biol.* 4(11):2467-2478), or a CUP1 promoter. Where desired, the subject vector comprise a promoter that is stronger than a native *E. Coli* Lac promoter.
- [00159] Non-limiting examples of operators for use in bacterial host cells include a lactose promoter operator (LacI repressor protein changes conformation when contacted with lactose, thereby preventing the LacI repressor protein from binding to the operator), a tryptophan promoter operator (when complexed with tryptophan, TrpR repressor protein has a conformation that binds the operator; in the absence of tryptophan, the TrpR repressor protein has a conformation that does not bind to the operator), and a tac promoter operator (see, for example, deBoer et al. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25.).
- [00160] The genes in the expression vector typically will also encode a ribosome binding site to direct translation (that is, synthesis) of any encoded mRNA gene product. For suitable ribosome binding sites for use in *Escherichia coli*, see Shine et al. (1975) *Nature* 254:34, and Steitz, in *Biological Regulation and Development: Gene Expression* (ed. R. F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, N.Y. Insertion of the ribosome binding site encoding nucleotide sequence 5'-AAAACA-3' upstream of a coding sequence facilitates efficient translation in a yeast host microorganism (Looman et al. (1993) *Nuc. Ac. Res.* 21:4268-4271; Yun et. al. (1996) *Mol. Microbiol.* 19:1225-1239).
- [00161] Other regulatory elements that may be used in an expression vector include transcription enhancer elements and transcription terminators. See, for example, Bitter et al. (1987) *Methods in Enzymology*, 153:516-544.
- [00162] An expression vector may be suitable for use in particular types of host microorganisms and not others. One of ordinary skill in the art, however, can readily determine through routine experimentation whether a particular expression vector is suited for a given host microorganism. For example, the expression vector can be introduced into the host organism, which is then monitored for viability and expression of any genes contained in the vector.
- [00163] The expression vector may also contain one or more selectable marker genes that, upon expression, confer one or more phenotypic traits useful for selecting or otherwise identifying host cells that carry the expression vector. Non-limiting examples of suitable selectable markers for eukaryotic cells include dihydrofolate reductase and neomycin resistance. Non-limiting examples of suitable selectable markers for prokaryotic cells include tetracycline, ampicillin, chloramphenicol, carbenicillin, and kanamycin resistance.

- [00164] For production of isoprenoid at an industrial scale, it may be impractical or too costly to use a selectable marker that requires the addition of an antibiotic to the fermentation media. Accordingly, some embodiments of the present invention employ host cells that do not require the use of an antibiotic resistance conferring selectable marker to ensure plasmid (expression vector) maintenance. In these embodiments of the present invention, the expression vector contains a plasmid maintenance system such as the 60-kb IncP (RK2) plasmid, optionally together with the RK2 plasmid replication and/or segregation system, to effect plasmid retention in the absence of antibiotic selection (see, for example, Sia et al. (1995) J. Bacteriol. 177:2789-97; Pansegrau et al. (1994) J. Mol. Biol. 239:623-63). A suitable plasmid maintenance system for this purpose is encoded by the *parDE* operon of RK2, which codes for a stable toxin and an unstable antitoxin. The antitoxin can inhibit the lethal action of the toxin by direct protein-protein interaction. Cells that lose the expression vector that harbors the *parDE* operon are quickly deprived of the unstable antitoxin, resulting in the stable toxin then causing cell death. The RK2 plasmid replication system is encoded by the *trfA* gene, which codes for a DNA replication protein. The RK2 plasmid segregation system is encoded by the *parCBA* operon, which codes for proteins that function to resolve plasmid multimers that may arise from DNA replication.
- [00165] The subject vectors can be introduced into a host cell stably or transiently by variety of established techniques. For example, one method involves a calcium chloride treatment wherein the expression vector is introduced via a calcium precipitate. Other salts, for example calcium phosphate, may also be used following a similar procedure. In addition, electroporation (that is, the application of current to increase the permeability of cells to nucleic acids) may be used. Other transformation methods include microinjection, DEAE dextran mediated transformation, and heat shock in the presence of lithium acetate. Lipid complexes, liposomes, and dendrimers may also be employed to transfect the host microorganism.
- [00166] Upon transformation, a variety of methods can be practiced to identify the host cells into which the subject vectors have been introduced. One exemplary selection method involves subculturing individual cells to form individual colonies, followed by testing for expression of the desired gene product. Another method entails selecting transformed host cells based upon phenotypic traits conferred through the expression of selectable marker genes contained within the expression vector. Those of ordinary skill can identify genetically modified host cells using these or other methods available in the art.
- [00167] The introduction of various pathway sequences of the invention into a host cell can be confirmed by methods such as PCR, Southern blot or Northern blot hybridization. For example, nucleic acids can be prepared from the resultant host cells, and the specific sequences of interest can be amplified by PCR using primers specific for the sequences of interest. The amplified product is subjected to agarose gel electrophoresis, polyacrylamide gel electrophoresis or capillary electrophoresis, followed by staining with ethidium bromide, SYBR Green solution or the like, or detection of DNA with a UV detection. Alternatively, nucleic acid probes specific for the sequences of interest can be employed in a hybridization reaction. The expression of a specific gene sequence can be ascertained by detecting the corresponding mRNA via reverse-transcription coupled PCR, Northern blot hybridization, or by immunoassays using antibodies reactive with the encoded gene product. Exemplary immunoassays include but are not limited to ELISA, radioimmunoassays, and sandwich immunoassays.
- [00168] The enzymatic activity of a given pathway enzyme can be assayed by a variety of methods known in the art. In general, the enzymatic activity can be ascertained by the formation of the product or conversion of a substrate of an enzymatic reaction that is under investigation. The reaction can take place in vitro or in

vivo. For example, the relative activity of HMG-CoA reductase and HMG-CoA synthase in a cell can be measured by the steady state level of HMG-CoA in a cell. HMG-CoA can be extracted by Trichloroacetic Acid (TCA), followed by analyzing the extracted material via Liquid Chromatography/Mass Spectrometry. The activity of mevalonate kinase can be demonstrated by the formation of mevalonate 5-phosphate. The relative activity of mevalonate kinase and HMG-CoA reductase can be measured by the steady state level of mevalonate, which can be determined by Gas Chromatography/Mass spectrometry. See e.g., WO05033287, which is incorporated herein by reference.

[00169] The yield of an isoprenoid via one or more metabolic pathways disclosed herein can be augmented by inhibiting reactions that divert intermediates from productive steps towards formation of the isoprenoid product. Inhibition of the unproductive reactions can be achieved by reducing the expression and/or activity of enzymes involved in one or more unproductive reactions. Such reactions include side reactions of the TCA cycle that lead to fatty acid biosynthesis, alanine biosynthesis, the aspartate superpathway, gluconeogenesis, heme biosynthesis, and/or glutamate biosynthesis, at a level that affects the overall yield of an isoprenoid production. Additionally, the conversion of acetyl-CoA to acetate via the action of phosphotransacetylase is another example of unproductive side reaction. Therefore, where desired, “knocking out” or “knocking down” the *pta* gene that encodes phosphotransacetylase may also be carried in order to increase the yield of isoprenoid production. Depending on the specific isoprenoid of interest, one skilled in the art may choose to target additional unproductive steps. For example, where carotenoid is the isoprenoid of choice, one may opt to “knock out” or “knock down” one or more genes selected from the group consisting of *gdhA*, *aceE*, *fdhF*, *yjiD*, *hnr* or *yjfP*, *ackA*, *appY*, *aspC*, *clp*, *clpP*, *clpXP*, *crcB*, *csdA*, *cyaA*, *evgS*, *fdhA*, *fdhD*, *feoB*, *fumA*, *glnE*, *glxR*, *gntK*, *hycI*, *lipB*, *lysU*, *modA*, *moeA*, *nadA*, *nuoC*, *nuoK*, *pflB*, *pitA*, *pst*, *pstC*, *pta*, *p-yjiD*, *sohA*, *stpA*, *yagR*, *yaiD*, *ybaS*, *ycfZ*, *ydeN*, *yebB*, *yedN*, *yfcC*, *ygiP*, *yibD*, *yjfP*, *yjhH*, or *yliE* genes, or any other genes alone or in combination, the inhibition of which would result in a higher yield of carotenoid as described in U.S. Patent Application 20060121558, which is incorporated herein by reference.

[00170] A variety of methods are available for knocking out or knocking down a gene of interest. For example, a reduced gene expression may be accomplished by deletion, mutation, and/or gene rearrangement. It can also be carried out with the use of antisense RNA, siRNA, miRNA, ribozymes, triple stranded DNA, and transcription and/or translation inhibitors. In addition, transposons can be employed to disrupt gene expression, for example, by inserting it between the promoter and the coding region, or between two adjacent genes to inactivate one or both genes.

High Yields of Isoprenoid Compounds

[00171] The present invention provides compositions and methods for a robust production of isoprenoids by the use of isopentenyl pyrophosphate pathway enzymes that are under the control of at least one heterologous regulator or fermentation conditions, either alone or in combination.

[00172] In one aspect, a method of producing an isoprenoid involves the steps of (a) obtaining a plurality of host cells that comprise an enzymatic pathway for making isopentenyl pyrophosphate wherein the all of the pathway enzymes are under control of at least one heterologous transcriptional regulator; and (b) culturing the host cells in a medium under conditions that are suboptimal as compared to conditions that would provide for a maximum specific growth rate for the host cells. In some embodiments, the pathway is the mevalonate pathway. In other embodiments, the pathway is the DXP pathway. In other embodiments, the at least one heterologous transcriptional regulatory sequence is inducible. In other embodiments, the

pathway enzymes are under control of a single transcriptional regulator. In other embodiments, the pathway enzymes are under control of multiple heterologous transcriptional regulators.

- [00173] In some embodiments, the pathway comprises a nucleic acid sequence encoding a mevalonate pathway enzyme from a prokaryote having an endogenous mevalonate pathway. Non-limiting examples of suitable prokaryotes include those from the genera: *Actinoplanes*; *Archaeoglobus*; *Bdellovibrio*; *Borrelia*; *Chloroflexus*; *Enterococcus*; *Lactobacillus*; *Listeria*; *Oceanobacillus*; *Paracoccus*; *Pseudomonas*; *Staphylococcus*; *Streptococcus*; *Streptomyces*; *Thermoplasma*; and *Vibrio*. Non-limiting examples of specific strains include: *Archaeoglobus fulgidus*; *Bdellovibrio bacteriovorus*; *Borrelia burgdorferi*; *Chloroflexus aurantiacus*; *Enterococcus faecalis*; *Enterococcus faecium*; *Lactobacillus johnsonii*; *Lactobacillus plantarum*; *Lactococcus lactis*; *Listeria innocua*; *Listeria monocytogenes*; *Oceanobacillus iheyensis*; *Paracoccus zeaxanthinifaciens*; *Pseudomonas mevalonii*; *Staphylococcus aureus*; *Staphylococcus epidermidis*; *Staphylococcus haemolyticus*; *Streptococcus agalactiae*; *Streptomyces griseolosporeus*; *Streptococcus mutans*; *Streptococcus pneumoniae*; *Streptococcus pyogenes*; *Thermoplasma acidophilum*; *Thermoplasma volcanium*; *Vibrio cholerae*; *Vibrio parahaemolyticus*; and *Vibrio vulnificus*;
- [00174] In another embodiment, the nucleic acid sequence encoding a mevalonate pathway enzyme is selected from acetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, and mevalonate kinase. In another embodiment, the nucleic acid sequence encoding a mevalonate pathway enzyme is selected from acetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, and mevalonate kinase and is from a prokaryote belonging to the genus *Enterococcus* or the genus *Pseudomonas* or the genus *Staphylococcus*. In another embodiment, the nucleic acid sequence encoding a mevalonate pathway enzyme is selected from acetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, and mevalonate kinase and is from *Enterococcus faecalis* or from *Staphylococcus aureus*.
- [00175] In another embodiment, nucleic acid sequence encoding a mevalonate pathway enzyme is a Class II HMG-CoA reductase. HMG-CoA reductases are generally classified into two classes, which are distinguishable based on sequence homology and/or enzymatic properties (see, for example, Hedl, et al., J. Bacteriology, 1927–1932, 2004, and Bochar, et al., Molec. Genet. Metab., 66, 122–127, 1999).
- [00176] Class II HMG-CoA reductases can be characterized, in part, by their low sensitivity to statins, including but not limited to Atorvastatin, Cerivastatin, Fluvastatin, Lovastatin, Pravastatin, Simvastatin, in some embodiments, the Class II HMG-CoA reductase exhibits a statin inhibition constant greater than about 1 micromolar, 10 micromolar, or 100 micromolar. In other embodiments, the Class II HMG-CoA reductase has an inhibition constant for Lovastatin greater than that of the Class I HMG-CoA by a factor of at least about 10, 100, 1000, or 10,000. In other embodiments, the Class II HMG-CoA reductase has an inhibition constant for Lovastatin greater than that of the Class I HMG-CoA isolated from a *Homo sapien* by a factor of at least about 10, 100, 1000, or 10,000. In other embodiments, the Class II HMG-CoA reductase is from a prokaryote. In other embodiments, the Class II HMG-CoA reductase is from archae bacteria.
- [00177] A prototypical Class II HMG-CoA reductase is derived from *Pseudomonas mevalonii*. Also encompassed in the invention are variant Class II HMG-CoA reductases exhibiting at least about 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 80%, 90%, or 95% identity as compared to the amino acid sequence of *P. mevalonii* HMG-CoA reductase. Further encompassed in the invention are variants having less than about 40%, 35%, 30%, 25%, 20%, or less, identity with an *H. Sapiens* HMG-CoA reductase. The identities of

amino acid sequences can be determined by the methods described in Bochar, et al., *Molec. Genet. Metab.*, 66, 122-127, 1999.

[00178] Non-limiting exemplary Class II HMG-CoA reductases include those derived from HMG-CoA reductases from: *Archaeoglobus fulgidus* (NC_000917); *Bdellovibrio bacteriovorus* (BX842650); *Borrelia burgdorferi* (AE001169); *Chloroflexus aurantiacus* (AJ299212); *Enterococcus faecalis* (AAO81155); *Enterococcus faecium* (AF290094); *Lactobacillus johnsonii* (AE017204); *Lactobacillus plantarum*; *Lactococcus lactis* (AE006387); *Listeria innocua* (CAC96053); *Listeria monocytogenes* (AE017324); *Oceanobacillus iheyensis* (NC_000917); *Paracoccus zeaxanthinifaciens* (AJ431696); *Pseudomonas mevalonii* (M24015); *Staphylococcus aureus* (AF290086); *Staphylococcus epidermidis* (AF290090); *Staphylococcus haemolyticus* (AF290088); *Streptococcus agalactiae* (CAD47046); *Streptomyces griseolosporeus* (AB037907); *Streptococcus mutans* (AAN58647); *Streptococcus pneumoniae* (AF290098); *Streptococcus pyogenes* (AF290096); *Thermoplasma acidophilum* (CAC11548); *Thermoplasma volcanium* (AL935253); *Vibrio cholerae* (AAF96622); *Vibrio parahaemolyticus* (BAC62311); and *Vibrio vulnificus* (AAO07090).

[00179] The fermentation methods described herein relate to modulating the specific growth rate of the host cells. Often represented by the parameter μ , the specific growth rate represents the rate of growth of cells per unit of biomass per unit time. The specific growth rate has the units of reciprocal time (1/t). The maximum specific growth rate for cells in a culture medium relates to the effect of substrate concentration on growth rate. Generally, cells will grow slowly at a low level of the substrate, and as the level of the substrate in the medium increases, so does the rate of cell growth. However, the rate of cell growth does not continue to rise indefinitely, and at high levels of substrate, a given increase in the amount of substrate will produce a smaller and smaller increase in the rate of cell growth. Therefore, the growth rate ultimately reaches a limit, which is often referred to as the maximum specific growth rate. A theoretical treatment of the relationship between growth rate in culture is well known to those skilled in the art, and is referred to as the Monod equation. See, for example, Pirt, *Principles of Microbe and Cell Cultivation*, Wiley, NY, 1975, pages 4-10. In this theoretical treatment, the maximum specific rate is an asymptotic limit that is never reached until an infinite level of substrate is reached. In practice, however, the maximum specific growth rate can be considered as being obtained when the conditions under investigation (e.g., a substrate level or temperature) support the fastest initial growth rate. For instance, in a fed-batch reactor, the initial condition where the nutrients are supplied in excess is treated as the conditions for the maximum growth rate. See, for example, Lee et al. (1996) *Trends Biotechnol.* 14: 98-105 and Korz et al. (1995) *J Biotechnology* 39:59-65. The conditions where a substrate is added to support the maximum specific growth rate is also sometimes referred to as unlimited growth. In addition, while the Monod equation describes the theoretical rate properties for a substrate that asymptotically approaches the maximum specific rate, for many substrates, rather than approaching the value as more substrate is added, a decrease in rate is seen at higher levels of substrate after a maximum rate is achieved, i.e., the maximum specific growth rate is achieved followed by a decrease in growth rate.

[00180] The maximum specific growth rate can also be applied with respect to temperature as well as to substrates. Generally, an organism will grow slowly at low temperatures, and will grow at a faster rate as the temperature increases up to a certain point, after which, the growth rate will decline. There will be a temperature at which the growth rate will be at a maximum level, this is the temperature at which the

maximum growth rate is achieved. We have found that the production of isoprenoids can be increased by lowering the temperature below the temperature that supports the maximum specific growth rate.

- [00181] The maximum specific growth rate can also be applied with respect to other additives to the fermentation than substrates. For instance, with respect to nutrients, vitamins, and minerals, there can be a low rate at low amounts of these components, the rate will get higher as the concentration of the component is increased, then, in some cases, at even higher concentrations of the components, the rate will decrease. The maximum specific growth rate is obtained where the concentration of the component supports the highest rate.
- [00182] The maximum specific growth rate for a cell in a medium is often determined at the initial stages of the fermentation before inhibition by end product or intermediates, cell crowding, or other factors contribute to slowing down the rate of growth. For example the maximum growth rate is often determined during the exponential phase of growth rather than at the lag phase, deceleration phase, or the stationary phase. The concept of maximum specific growth rate can also be applied at later stages of the fermentation by taking into account the appropriate variables.
- [00183] Accordingly, in some embodiments, host cells are cultured under conditions such that growth is less than about 90% of the maximum specific growth rate. In other embodiments, the host cells are cultured under conditions such that growth is less than about 80%, 75%, 60%, 50%, 40%, 30%, 25%, 20%, 10%, 5%, or 1%, or less, of the maximum specific growth rate.
- [00184] In other embodiments, the host cells are cultured at a medium temperature is at least about 2°C, 4°C, 5°C, 6°C, 8°C, 10°C, 15°C, or 20°C below the temperature that would provide for the maximum specific growth rate. By lowering the temperature growth is reduced, which in turn, reduces the formation of toxic byproducts in the medium and the generation of metabolic heat. Lowering culture temperature also reduces cellular oxygen demand which enables higher cell-densities to be obtained.
- [00185] The temperature at which at which the maximum specific growth rate of a host cell can be achieved will depend on the type of host cell selected. This can be ascertained by growing the host cells under various temperatures over a defined period of time to derive the relevant growth curves. The temperature that supports the maximum specific growth rate can be determined by comparing the slopes of growth in the respective curves. In the case of *E. Coli*, the temperature for maximum specific growth rate is about 37°C. Accordingly, if *E. Coli* is the host cell to be used for fermentative production of isoprenoid, the fermentation temperature is below 37°C. If *S. cerevisiae* is employed, the temperature for maximum specific growth rate is about 30°C. Accordingly, if *S. cerevisiae* is the host cell to be used for fermentative production of isoprenoid, the fermentation temperature is below 30°C. Typically a desired temperature is about 2 °C, 4 °C, 5 °C, 6, 8, 10 °C, 15 °C, and 20 °C below the temperature at which the maximum specific growth rate of the host cell can be achieved.
- [00186] In other embodiments, the host cells are cultured in a fermentation medium comprises a carbon source present in an amount that is lower than that which would provide for a maximum specific growth rate. In certain embodiments, the host cells are cultured in a medium where the carbon source is maintained at a level to provide for less than about 90%, 80%, 75%, 60%, 50%, 40%, 30%, 25%, 20%, 10%, 5%, 1%, or less, of the maximum specific growth rate. Any carbon-containing sources that are digestible by the microorganism can be used. Non-limiting examples include carbohydrates such as monosaccharides, oligosaccharides and polysaccharides, organic acids such as acetic acid, propionic acid; and alcohols such as ethanol and propanol, and polyols such as glycerol.

- [00187] In some embodiments, the carbon sources comprise primarily monosaccharides or oligosaccharides. In other embodiments, the carbon source consists essentially of monosaccharides and disaccharides. In still other embodiments, the carbon source is essentially free of cellulose.
- [00188] Monosaccharides are the simple sugars that serve as building blocks for carbohydrates. They are classified based on their backbone of carbon (C) atoms: trioses have three carbon atoms, tetroses four, pentoses five, hexoses six, and heptoses seven. The carbon atoms are bonded to hydrogen atoms ($-H$), hydroxyl groups ($-OH$), and carbonyl groups ($-C=O$), whose combinations, order, and configurations allow a large number of stereoisomers to exist. Pentoses include xylose, found in woody materials; arabinose, found in gums from conifers; ribose, a component of RNA and several vitamins, and deoxyribose, a component of DNA. Exemplary hexoses include glucose, galactose, and fructose. Monosaccharides combine with each other and other groups to form a variety of disaccharides, and oligosaccharides. An oligosaccharide is a saccharide polymer containing a small number (typically three to ten) of simple sugars. They are generally found either *O*- or *N*-linked to compatible amino acid side chains in proteins or lipid moieties. A preferred oligosaccharide for use in the present fermentation reaction is disaccharide, including for example, sucrose, or trisaccharide such as raffinose.
- [00189] Where it is desired to have cellulose, glycan, starch, or other polysaccharides as the ultimate carbon source, these polysaccharides can be first converted into monosaccharides and oligosaccharides by chemical means or by enzymatic methods. For instance, cellulose can be converted into glucose by the enzyme cellulase. Accordingly, if polysaccharides such as cellulose found in the biomass (including e.g., canola, alfalfa, rice, rye, sorghum, sunflower, wheat, soybean, tobacco, potato, peanut, cotton, sweet potato, cassava, coffee, coconut, citrus trees, cocoa, tea, fruits such as, banana, fig, pineapple, guava, mango, oats, barley, vegetables, ornamentals, or conifers) is used as the ultimate carbon source, it can be digested by cellulase to generate simpler sugars for use in conjunction with the fermentation procedure of the present invention. In certain embodiments, after the breakdown of the polysaccharide, the monosaccharide and/or oligosaccharide constitute at least about 50% by weight of the carbon source as determined at the beginning of the fermentation. In other embodiments, the monosaccharide and/or oligosaccharide constitute at least about 80% or even 90% by weight of the carbon source as determined at the beginning of the fermentation, such that the fermentation medium is essentially free of cellulose.
- [00190] In other embodiments, the host cells are cultured in a fermentation medium comprises a nitrogen source present in an amount that is lower than that which would provide for a maximum specific growth rate. While not being bound by any particular theory, it is known that changing the levels of components such as nitrogen that are available to a cell can change the relative flux through the various chemical pathways within the cell. We have found that by restricting the level of nitrogen available to the microorganism, the amount of isoprenoid such as amorphaadiene produced by the microorganism is increased. Exemplary levels of nitrogen of the present invention include in an amount that would support about 90%, 80%, 75%, 60%, 50%, 40%, 30%, 25%, 20%, 10%, 5%, 1%, or less, of the maximum specific growth rate.
- [00191] The restriction of nitrogen can be implemented in stages. In some embodiments, nitrogen in the form of ammonia is provided in the beginning of the fermentation to support initial growth, but subsequent additions to the fermentation are free of nitrogen, or are free of nitrogen save for that level of ammonia needed to maintain the pH of the fermentation at 7 with an ammonia solution. For the bulk of the fermentation, the level of nitrogen is maintained at a level which is less than the amount which would support the maximum specific growth rate. The amounts can be for example amounts which would support

at least about 90%, 80%, 75%, 60%, 50%, 40%, 30%, 25%, 20%, 10%, 5%, or 1%, or less, of the maximum specific growth rate. A fermentation of the present invention could have an initial nitrogen level above 10 mM as measured in the fermentation medium, to support an initial growth, and a subsequent a nitrogen level below 50mM, 40mM, 30mM, 20mM, 10mM, or 4 mM in the fermentation medium.

[00192] Sources of assimilable nitrogen that can be used in a suitable fermentation reaction mixture include, but are not limited to, simple nitrogen sources, organic nitrogen sources, and complex nitrogen sources. Such nitrogen sources include anhydrous ammonia, ammonium salts of inorganic or organic acids such as ammonium chloride, ammonium sulfate, ammonium acetate, ammonium phosphate, other nitrogen-containing compounds and substances of animal, vegetable, and/or microbial origin. Amino acids can also be used as the nitrogen source, including leucine, isoleucine or valine, or a mixture thereof.

[00193] Any known method for providing a substrate to a fermentation reaction may be used to maintain the substrate level below the level which would provide for the maximum specific growth rate. Illustrative examples include the batch method where all the substrate for the fermentation is added in the beginning of the fermentation reaction; the continuous feed method; and the variable feed rate method, where, for instance, an increasing amount of substrate is provided as the fermentation proceeds in order to support the increased concentration of cells in the medium. Combinations of these three methods are often employed. For instance, it is common to have a certain amount of substrate present initially in the fermentation, to allow the microorganisms to deplete this initial amount of substrate, then to subsequently add substrate either continuously or to add it variably after the initial amount of substrate is utilized. It is an aspect of this invention to provide an initial amount of substrate to the cells in the fermentation medium, which may be present at a relatively high level, to allow the host cells to substantially use up the initial substrate, then to subsequently provide substrate to the host cells at a level that is suboptimal as compared to the amount that would support the maximum growth rate by either a continuous or variable feed rate. It will be appreciated by those of skill in the art that since the cells may be growing at an exponential rate, it can be advantageous to vary the feed rate at an exponential rate in order to keep the amount of substrate relative to the level of host cells constant. Thus in certain embodiments, substrates are added in an exponentially increasing manner, but yet at a level which is lower than the level which would provide the maximum specific growth rate.

[00194] In some embodiments, the fermentation reaction is given a reduced feed of carbon source relative to the carbon feed which would provide for the maximum specific growth rate. While not being bound by a particular theory, it is known that changing the levels of nutrients available to a cell will change the relative flux through the various chemical pathways within the cell. For example, some enzymes are inducible, and will only be active when certain nutrients are present. We have observed that lowering the carbon source feed rate to a microorganism can improve the amount of isoprenoid produced in the fermentation. In practice, the carbon source can be supplied initially in an amount sufficient to support an initial growth of the host cells until such initial carbon source is substantially depleted, after which the carbon source is added at an exponential rate, but at a rate which is below that which would support the maximum specific growth of the host cells. For example, in a method of the present invention, the carbon source is added in an amount that would support about 90%, 80%, 75%, 60%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, or less, of the maximum specific growth rate.

[00195] In other embodiments, the fermentation reaction is given an initial bolus of carbon source sufficient to grow at or near the maximum specific growth rate (unlimited growth) followed by a reduced feed rate at a level

below that required to support the maximum specific growth rate, for the remainder of the fermentation. In some cases, the point at which the reduced carbon source feed rate is implemented is the point at which a predetermined feed rate is achieved. In certain embodiments, the microorganism is provided with enough carbon source to grow exponentially to a feed rate of about 15 g/L/hr, after which the feed rate is reduced to 5.7 g/L/hr and held constant at that rate for the remainder of the fermentation.

- [00196] In certain embodiments, one or more of the heterologous mevalonate or DXP pathway enzymes is inducible and induced after the carbon source feed rate has been reduced to a level below that required for maximum specific growth. For example, where the engineered microorganism has an inducible promoter, the fermentation is first run by adding carbon source to achieve an exponential growth, but at a level which is below that to support maximum specific growth, then the carbon source feed rate is reduced to an even lower level for the remainder of the fermentation, and the inducer added after the carbon source feed rate is reduced. In some embodiments, the microorganisms are induced with isopropylthio-beta-D-galactoside (IPTG) after the reduced carbon source feed is initiated.
- [00197] In other embodiments, the fermentation reaction is performed in a manner that avoids the build up of toxic substances that decrease cell growth rates. For example, it is known that when too much glucose is added to the medium, toxic products such as acetate can build up in the organism. See, for example, Kortz et al. (1995) *J. Biotechnol.* 39: 59-65. Thus by providing a high level of carbon source, at or approaching the amount which would support a maximum growth rate (unlimited growth), the initial growth of the cells may be higher, but the growth becomes arrested due to the accumulation of toxic substances. The level at which the carbon source is added below the level where the toxic products do not accumulate is referred to as the critical level or the inhibitory threshold. Thus in certain embodiments, the fermentation reaction is performed such that the carbon source is kept below the critical level for the build up of toxic substances. Those skilled in the art will appreciate that the critical concentration of substrates will vary with the strain and the medium which is used.
- [00198] An effective fermentation reaction mixture can contain other compounds such as inorganic salts, vitamins, trace metals, or growth promoters. Such other compounds can also be present in carbon, nitrogen or mineral sources in the effective reaction mixture or can be added specifically to the reaction mixture. One embodiment of the invention involves providing these compounds at levels that are suboptimal as compared to that would support the maximum growth rate of the host cells in order to increase isoprenoid production.
- [00199] The fermentation reaction mixture can also contain a suitable phosphate source. Such phosphate sources include both inorganic and organic phosphate sources. Non-limiting examples of phosphate sources include, but are not limited to, phosphate salts such as mono or dibasic sodium and potassium phosphates, ammonium phosphate, polyphosphate, and mixtures thereof. A suitable fermentation reaction mixture can also include a source of magnesium. In some embodiments, the magnesium is in the form of a physiologically acceptable salt, such as magnesium sulfate heptahydrate, although other magnesium sources in concentrations that contribute similar amounts of magnesium can be used. Further, in some instances it may be desirable to allow the fermentation reaction mixture to become depleted of a magnesium source during fermentation. In some embodiments, the phosphorous source is provided in an amount that is suboptimal as compared to that would support a maximum specific growth rate.
- [00200] The fermentation reaction mixture can also include a biologically acceptable chelating agent, such as the dihydrate of trisodium citrate and ethylenediaminetetraacetic acid. The fermentation reaction mixture can

also initially include a biologically acceptable acid or base to maintain the desired pH of the fermentation reaction mixture. Biologically acceptable acids include, but are not limited to, hydrochloric acid, sulfuric acid, nitric acid, phosphoric acid and mixtures thereof. Biologically acceptable bases include, but are not limited to, ammonium hydroxide, sodium hydroxide, potassium hydroxide and mixtures thereof.

[00201] The fermentation reaction mixture can also include a biologically acceptable calcium source, including, but not limited to, calcium chloride. The fermentation reaction mixture can also include sodium chloride. The fermentation reaction mixture can also include trace metals. Such trace metals can be added to the fermentation reaction mixture as a stock solution that, for convenience, can be prepared separately from the rest of the fermentation reaction mixture. A suitable trace metals solution can include, but is not limited to sodium selenate; ferrous sulfate; heptahydrate; cupric sulfate, pentahydrate; zinc sulfate, heptahydrate; sodium molybdate, dihydrate; cobaltous chloride; selenium or chromium solution; hexahydrate; and manganous sulfate monohydrate. Hydrochloric acid may be added to the stock solution to keep the trace metal salts in solution.

[00202] If a pathway intermediate or a compound that can be converted to a pathway intermediate is added to the fermentation medium, the intermediate or compound is typically present in an excess amount.

[00203] Fermentation can be conducted under anaerobic (deficient in oxygen) or aerobic (oxygenated) conditions. Under aerobic conditions, microorganisms can break down sugars to end products such as CO₂ and H₂O. Under anaerobic conditions, the host cells utilize an alternative pathway to produce CO₂ and ethanol. Fermentation can also be used to refer to the bulk growth of microorganisms on a growth medium where no distinction is made between aerobic and anaerobic metabolism. In general, aerobic fermentation is carried out for production of isoprenoids.

[00204] The fermentations of the present invention can be carried out in a batch, a fed-batch, or a continuous process. A batch process is typically a closed process where all of the raw materials are added at the beginning of the fermentation. A fed-batch process is typically a closed process where the carbon source and/or other substrates are added in increments throughout the process. A fed-batch process allows for greater control of the medium and the growth of the microorganisms. A continuous process can be considered an open system where medium is continuously added and product is simultaneously removed. Processes in between these types can also be used. For instance, in one embodiment, the fermentation is begun as a fed-batch process, and an organic layer, such as dodecane is placed in contact with the fermentation medium while the fermentation process continues. Isoprenoids, which typically have a higher solubility in the organic medium than in the aqueous fermentation medium are extracted out of the fermentation medium into the organic layer. Where the isoprenoids are produced in excess of the saturation point and form a layer separable from the medium, then simple separation by way of draining or sucking the distinct phase layer can be carried out. This process has characteristics of both a fed-batch process and a continuous process, because of the removal of product from the medium and the fermentation progresses. The fed-batch and continuous processes allow for the control of the addition of fermentation components during the fermentation process. A fed-batch, continuous, or combination of these processes is usually preferred in carrying out the invention. These processes allow for greater control of the rate of addition of feed and other fermentation components as a function of time. The removal of product during fermentation can be beneficial, especially where the accumulated product leads to inhibition of the production pathways.

- [00205] The amount of microorganism per liter of fermentation, or the density of microorganism, can be measured by measuring the weight of microorganism isolated from a given volume of the fermentation medium. A common measure is the dry weight of cells per liter of fermentation medium. Another method which can be used to monitor the fermentation while it is progressing is by a measurement of the optical density of the medium. A common method is to measure the optical density at a wavelength of 600 nm, referred to the OD₆₀₀, or the OD. The OD can be correlated to a the density of a specific type of organism within a specific medium, but the specific relationship between OD and amount of microorganism per volume will not generally be applicable across all types of organisms in all types of media. A calibration curve can be created by measuring the OD and the dry cell weight over a range of cell densities. In some cases, these correlations can be used in different fermentation of the same or similar microorganisms in the same or similar media.
- [00206] In another aspect, the present invention provides a method comprising the steps of (i) performing a fermentation reaction comprising a fermentation medium and a plurality of genetically modified host cells that produce the isoprenoid under conditions such that (a) the fermentation medium is kept at a temperature lower than that which would provide for a maximum specific growth rate of said host cells; (b) the fermentation medium comprises a carbon source present in an amount that is lower than that which would provide for a maximum specific growth rate of the host cells; and/or (c) the fermentation medium comprises a nitrogen source present in an amount that is lower than that which would provide for a maximum specific growth rate of the host cells; (ii) recovering the isoprenoid produced under one or more conditions set forth in (a) through (c). In one embodiment, the fermentation reaction is run under condition (a). In another embodiment, the fermentation reaction is run under conditions (a) and (b). In yet another embodiment, the fermentation reaction is run under conditions of (a), (b), and (c), or in any other combinations thereof.
- [00207] Using the methods described herein, the host cells produce more than about 10 grams of isoprenoid per liter of fermentation reaction mixture (10g/L). In other embodiments, more than about 15g/L, more than about 20g/L, more than 25g/L is produced, or more than about 30g/L of isoprenoid is produced.
- [00208] In another embodiment, the host cells produce more than about 50 milligrams of isoprenoid per gram of dry host cells (50 milligrams per gram dry cell weight) is produced. In other embodiments, more than about 100 milligrams per gram dry cell weight, more than about 150 milligrams per gram dry cell weight, more than about 200 milligrams per gram dry cell weight, more than about 250 milligrams per gram dry cell weight, more than about 500 milligrams per gram dry cell weight, more than about 750 milligrams per gram dry cell weight, or more than about 1000 milligrams per gram dry cell weight of isoprenoid is produced.
- [00209] In other embodiments, the production level, whether it is in grams per liter or milligrams per gram dry cell weight is achieved in less than about 150 hours, preferably less than about 96 hours, or even less than about 72 hours.
- [00210] Non-limiting examples of suitable isoprenoids include: hemiterpenes (derived from 1 isoprene unit) such as isoprene; monoterpenes (derived from 2 isoprene units) such as myrcene; sesquiterpenes (derived from 3 isoprene units) such as amorpho-4,11-diene; diterpenes (derived from four isoprene units) such as taxadiene; triterpenes (derived from 6 isoprene units) squalene; tetraterpenes (derived from 8 isoprenoids) β -carotene; and polyterpenes (derived from more than 8 isoprene units) such as polyisoprene. In some

embodiments, the isoprenoid is not a carotenoid. In other embodiments, the isoprenoid is a C₅-C₂₀ isoprenoid.

[00211] Although the invention has been described in conjunction with specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

EXAMPLES

[00212] The practice of the present invention can employ, unless otherwise indicated, conventional techniques of the biosynthetic industry and the like, which are within the skill of the art. To the extent such techniques are not described fully herein, one can find ample reference to them in the scientific literature.

[00213] In the following examples, efforts have been made to ensure accuracy with respect to numbers used (for example, amounts, temperature, and so on), but variation and deviation can be accommodated, and in the event a clerical error in the numbers reported herein exists, one of ordinary skill in the arts to which this invention pertains can deduce the correct amount in view of the remaining disclosure herein. Unless indicated otherwise, temperature is reported in degrees Celsius, and pressure is at or near atmospheric pressure at sea level. All reagents, unless otherwise indicated, were obtained commercially. The following examples are intended for illustrative purposes only and do not limit in any way the scope of the present invention.

Example 1

[00214] This example describes methods for making expression plasmids that encode for enzymes including enzymes of the MEV pathway from *Saccharomyces cerevisiae* organized in operons.

[00215] Expression plasmid pMevT was generated by inserting the MevT operon (SEQ ID NO: 1) into the pBAD33 vector. The MevT operon encodes the set of MEV pathway enzymes that together transform the ubiquitous precursor acetyl-CoA to (R)-mevalonate, namely acetoacetyl-CoA thiolase, HMG-CoA synthase, and HMG-CoA reductase. The MevT operon was generated by PCR amplifying from *Escherichia coli* genomic DNA the coding sequence of the *atoB* gene (GenBank accession number NC_000913 REGION: 2324131..2325315) (encodes an acetoacetyl-CoA thiolase), from *Saccharomyces cerevisiae* genomic DNA the coding sequence of the *ERG13* gene (GenBank accession number X96617, REGION: 220..1695) (encodes a HMG-CoA synthase), and from *Saccharomyces cerevisiae* genomic DNA a segment of the coding region of the *HMG1* gene (GenBank accession number M22002, REGION: 1660..3165) (encodes a truncated HMG-CoA reductase (tHMGR)). The upstream PCR primer used for the amplification of the *HMG1* gene fragment included an artificial start codon. The amplified fragments were spliced together using overlap extensions (SOEing), during which process ribosome binding sites were introduced after the *atoB* and the *ERG13* coding sequences. After the addition of 3' A overhangs, the MevT operon was ligated into the TA cloning vector pCR4 (Invitrogen, Carlsbad, CA), and sequenced to ensure accuracy. The MevT operon was subsequently ligated into the *XmaI* *PstI* restriction enzyme site of vector pBAD33 (Guzman et al. (1995) *J. Bacteriol.* 177(14): 4121-4130). To place the operon under the control of the P_{Lac} promoter, the *araC-P_{BAD}NsiI-XmaI* fragment of pBAD33 was replaced with the *NsiI-XmaI* fragment of pBBR1MCS, yielding expression plasmid pMevT (see U.S. Patent Number 7,192,751).

[00216] Expression plasmid pAM36-MevT66 was generated by inserting the MevT66 operon into the pAM36 vector. Vector pAM36 was generated by inserting an oligonucleotide cassette containing *AscI-SfiI-AsiSI-XhoI-PacI-FsiI-PmeI* restriction enzyme sites into the pACYC184 vector (GenBank

accession number XO6403), and by removing the *tet* resistance gene in pACYC184. The MevT66 operon was synthetically generated using the nucleotide sequence SEQ ID NO: 1 as a template, which comprises the *atoB* gene from *Escherichia coli* (GenBank accession number NC_000913 REGION: 2324131..2325315), the *ERG13* gene from *Saccharomyces cerevisiae* (GenBank accession number X96617, REGION: 220..1695), and a truncated version of the *HMG1* gene from *Saccharomyces cerevisiae* (GenBank accession number M22002, REGION: 1777..3285), all three sequences being codon-optimized for expression in *Escherichia coli*. The synthetically generated MevT66 operon was flanked by a 5' *EcoRI* restriction enzyme site and a 3' *Hind III* restriction enzyme site, and could thus be cloned into compatible restriction enzyme sites of a cloning vector such as a standard pUC or pACYC origin vector. From this construct, the MevT66 operon was PCR amplified with flanking *SfiI* and *AsiSI* restriction enzyme sites, the amplified DNA fragment was digested to completion using *SfiI* and *AsiSI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 4.2 kb DNA fragment was gel extracted using a Qiagen gel purification kit (Valencia, CA), and the isolated DNA fragment was ligated into the *SfiI* *AsiSI* restriction enzyme site of the pAM36 vector, yielding expression plasmid pAM36-MevT66.

[00217] Expression plasmid pAM25 was generated by inserting the MevT66 operon into the pAM29 vector. Vector pAM29 was created by assembling the p15A origin of replication and *kan* resistance gene from pZS24-MCS1 (Lutz and Bujard (1997) *Nucl Acids Res.* **25**:1203-1210) with an oligonucleotide-generated *lacUV5* promoter. The DNA synthesis construct comprising the MevT66 operon (see above) was digested to completion using *EcoRI* and *Hind III* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the 4.2 kb DNA fragment was gel extracted, and the isolated DNA fragment was ligated into the *EcoRI* *HindIII* restriction enzyme site of pAM29, yielding expression plasmid pAM25.

[00218] Expression plasmid pMevB-Cm was generated by inserting the MevB operon into the pBBR1MCS-1 vector. The MevB operon encodes the set of enzymes that together convert (R)-mevalonate to IPP, namely mevalonate kinase, phosphomevalonate kinase, and mevalonate pyrophosphate carboxylase. The MevB operon was generated by PCR amplifying from *Saccharomyces cerevisiae* genomic DNA the coding sequences of the *ERG12* gene (GenBank accession number X55875, REGION: 580..1911) (encodes a mevalonate kinase), the *ERG8* gene (GenBank accession number Z49939, REGION: 3363..4718) (encodes a phosphomevalonate kinase), and the *MVDI* gene (GenBank accession number X97557, REGION: 544..1734) (encodes a mevalonate pyrophosphate carboxylase), and by splicing the PCR fragments together using overlap extensions (SOEing). By choosing appropriate primer sequences, the stop codons of *ERG12* and *ERG8* were changed from TAA to TAG during amplification to introduce ribosome binding sites. After the addition of 3' A overhangs, the MevB operon was ligated into the TA cloning vector pCR4 (Invitrogen, Carlsbad, CA). The MevB operon was excised by digesting the cloning construct to completion using *PstI* restriction enzyme, resolving the reaction mixture by gel electrophoresis, gel extracting the 4.2 kb DNA fragment, and ligating the isolated DNA fragment into the *PstI* restriction enzyme site of vector pBBR1MCS-1 (Kovach *et al.*, *Gene* 166(1): 175-176 (1995)), yielding expression plasmid pMevB-Cm.

[00219] Expression plasmid pMBI was generated by inserting the MBI operon into the pBBR1MCS-3 vector. The MBI operon encodes the same enzymes as the MevB operon, as well as an isopentenyl pyrophosphatase isomerase that catalyzes the conversion of IPP to DMAPP. The MBI operon was generated by PCR amplifying from *Escherichia coli* genomic DNA the coding sequence of the *idi* gene (GenBank accession number AF119715) using primers that contained an *XmaI* restriction enzyme site at their 5' ends, digesting the amplified DNA fragment to completion using *XmaI* restriction enzyme, resolving the reaction mixture

by gel electrophoresis, gel extracting the 0.5 kb fragment, and ligating the isolated DNA fragment into the *XmaI* restriction enzyme site of expression plasmid pMevB-Cm, thereby placing *idi* at the 3' end of the MevB operon. The MBI operon was subcloned into the *Sall* and *SacI* restriction enzyme sites of vector pBBR1MCS-3 (Kovach *et al.*, *Gene* 166(1): 175-176 (1995)), yielding expression plasmid pMBI (see U.S. Patent Number 7,192,751).

[00220] Expression plasmid pMBIS was generated by inserting the *ispA* gene into pMBI. The *ispA* gene encodes a farnesyl pyrophosphate synthase that catalyzes the condensation of two molecules of IPP with one molecule of DMAPP to make farnesyl pyrophosphate (FPP). The coding sequence of the *ispA* gene (GenBank accession number D00694, REGION: 484..1383) was PCR amplified from *Escherichia coli* genomic DNA using a forward primer with a *SacII* restriction enzyme site and a reverse primer with a *SacI* restriction enzyme site. The amplified PCR product was digested to completion with *SacII* and *SacI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, and the 0.9 kb DNA fragment was gel extracted. The isolated DNA fragment was ligated into the *SacII SacI* restriction enzyme site of pMBI, thereby placing the *ispA* gene 3' of *idi* and the MevB operon, and yielding expression plasmid pMBIS (see U.S. Patent Number 7,192,751).

[00221] Expression plasmid pMBIS-gpps was derived from expression plasmid pMBIS by replacing the *ispA* coding sequence with a nucleotide sequence encoding a geranyl diphosphate synthase ("gpps"). A DNA fragment comprising a nucleotide sequence encoding the geranyl diphosphate synthase was generated synthetically using the coding sequence of the *gpps* gene of *Arabidopsis thaliana* (GenBank accession number Y17376, REGION: 52..1320), codon-optimized for expression in *Escherichia coli*, as a template. The nucleotide sequence was flanked by a leader *SacII* restriction enzyme site and a terminal *SacI* restriction enzyme site, and can be cloned into compatible restriction enzyme sites of a cloning vector such as a standard pUC or pACYC origin vector. The synthetically generated geranyl diphosphate synthase sequence was isolated by digesting the DNA synthesis construct to completion using *SacII* and *SacI* restriction enzymes, resolving the reaction mixture by gel electrophoresis, gel extracting the approximately 1.3 kb DNA fragment, and ligating the isolated DNA fragment into the *SacII SacI* restriction enzyme site of expression plasmid pMBIS, yielding expression plasmid pMBIS-gpps (see Figure 3 for a plasmid map).

[00222] Expression plasmid pAM45 was generated by inserting the MBIS operon into pAM36-MevT66 and adding *lacUV5* promoters in front of the two operons. The MBIS operon was PCR amplified from pMBIS using primers comprising a 5' *XhoI* restriction enzyme site and a 3' *PacI* restriction enzyme site. The amplified PCR product was digested to completion using *XhoI* and *PacI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the 5.4 kb DNA fragment was gel extracted, and the isolated DNA fragment was ligated into the *XhoI PacI* restriction enzyme site of pAM36-MevT66, yielding plasmid pAM43. A DNA fragment comprising a nucleotide sequence encoding the *lacUV5* promoter was synthesized from oligonucleotides and sub-cloned into the *AscI SfiI* and *AsiSI XhoI* restriction enzyme sites of pAM43, yielding expression plasmid pAM45.

Example 2

[00223] This example describes methods for making expression vectors encoding enzymes including enzymes of the MEV pathway from *Staphylococcus aureus* organized in operons.

[00224] Expression plasmid pAM41 was derived from expression plasmid pAM25 by replacing the coding sequence of the *HMG1* gene, which encodes the *Saccharomyces cerevisiae* HMG-CoA reductase, with the coding sequence of the *mvaA* gene, which encodes the *Staphylococcus aureus* HMG-CoA reductase

(GenBank accession number BA000017, REGION: 2688925..2687648). The coding sequence of the *mvaA* gene was PCR amplified from *Staphylococcus aureus* subsp. *aureus* (ATCC 70069) genomic DNA using primers 4-49 *mvaA* *SpeI* (SEQ ID NO: 2) and 4-49 *mvaA* *XbaI* (SEQ ID NO: 3), the amplified DNA fragment was digested to completion using *SpeI* restriction enzyme, the reaction mixture was resolved by gel electrophoresis, and the approximately 1.3 kb DNA fragment was gel extracted. The *HMG1* coding sequence was removed from pAM25 by digesting the plasmid to completion using *HindIII* restriction enzyme. The terminal overhangs of the resulting linear DNA fragment were blunted using T4 DNA polymerase. The DNA fragment was then partially digested using *SpeI* restriction enzyme, the reaction mixture was resolved by gel electrophoresis, and the 4.8 kb DNA fragment was gel extracted. The isolated DNA fragment was ligated with the *SpeI*-digested *mvaA* PCR product, yielding expression plasmid pAM41. The nucleotide sequence of the *atoB*(opt):*ERG13*(opt):*mvaA* operon contained in pAM41 is SEQ ID NO: 41.

[00225] Expression plasmid pAM52 was derived from expression plasmid pAM41 by replacing the coding sequence of the *ERG13* gene, which encodes the *Saccharomyces cerevisiae* HMG-CoA synthase, with the coding sequence of the *mvaS* gene, which encodes the *Staphylococcus aureus* HMG-CoA synthase (GenBank accession number BA000017, REGION: 2689180..2690346). *ERG13* is also known as HMGS or HMG-CoA synthase. The coding sequence of the *mvaS* gene was PCR amplified from *Staphylococcus aureus* subsp. *aureus* (ATCC 70069) genomic DNA using primers HMGS 5' *Sa mvaS*-S (SEQ ID NO: 4) and HMGS 3' *Sa mvaS*-AS (SEQ ID NO: 5), and the amplified DNA fragment was used as a PCR primer to replace the coding sequence of the *HMG1* gene in pAM41 according to the method of Geiser *et al.* (*BioTechniques* 31:88-92 (2001)), yielding expression plasmid pAM52. The nucleotide sequence of the *atoB*(opt):*mvaS*:*mvaA* operon contained in pAM52 is SEQ ID NO: 42.

[00226] Expression plasmid pAM97 was derived from expression plasmid pAM45 by replacing the *MevT66* operon with the *(atoB*(opt):*mvaS*:*mvaA*) operon of expression plasmid pAM52. Expression plasmid pAM45 was digested to completion using *AsiSI* and *SfiI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, and the 8.3 kb DNA fragment lacking the *MevT66* operon was gel extracted. The *(atoB*(opt):*mvaS*:*mvaA*) operon of pAM52 was PCR amplified using primers 19-25 *atoB* *SfiI*-S (SEQ ID NO: 6) and 19-25 *mvaA*-*AsiSI*-AS (SEQ ID NO: 7), the PCR product was digested to completion using *SfiI* and *AsiSI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the 3.7 kb DNA fragment was gel extracted, and the isolated DNA fragment was ligated into the *AsiSI* *SfiI* restriction enzyme site of expression plasmid pAM45, yielding expression plasmid pAM97.

[00227] Expression plasmid pAM97-MBI was derived from expression plasmid pAM97 and pAM45 by replacing the MBIS operon of pAM97 with the MBI operon of pAM45. The MBI operon was PCR amplified from pAM45 using primers 9-70C (SEQ ID NO: 8) and 26-39B (SEQ ID NO: 9), the reaction mixture was resolved by gel electrophoresis, the 4.5 kb DNA fragment was gel extracted, and the isolated DNA fragment was digested to completion using *SacI* and *XhoI* restriction enzymes. Expression plasmid pAM97 was digested to completion using *SacI* and *XhoI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the 7.6 kb fragment was gel extracted, and the isolated DNA fragment was ligated with the MBI operon PCR product, yielding expression plasmid pAM97-MBI.

[00228] Expression plasmid pAM97-MevB was derived from expression plasmid pAM97 and pAM45 by replacing the MBIS operon of pAM97 with the *MevB* operon of pAM45. The *MevB* operon was PCR amplified from

pAM43 using primers 9-70C (SEQ ID NO: 8) and 26-39A (SEQ ID NO: 10), the reaction mixture was resolved by gel electrophoresis, the 3.9 kb DNA fragment was gel extracted, and the isolated DNA fragment was digested to completion using *SacI* and *XhoI* restriction enzymes. Expression plasmid pAM97 was digested to completion using *SacI* and *XhoI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the 7.6 kb fragment was gel extracted, and the isolated DNA fragment was ligated with the *MevB* operon PCR product, yielding expression plasmid pAM97-MevB.

- [00229] Expression plasmid pAM128 was generated by inserting the (*atoB*(opt):*mvaS*:*mvaA*) and MBIS operons of expression plasmid pAM97 into a vector that comprises the RK2 plasmid replication, segregation, and maintenance system, which obviates the continuous need for antibiotic selection of host cell transformants. The RK2 plasmid was digested to completion using *PstI* restriction enzyme, the reaction mixture was resolved by gel electrophoresis, the approximately 6.3 kb DNA fragment containing the entire *par* locus was gel extracted, and the isolated DNA fragment was subcloned into the *PstI* restriction enzyme site of the mini RK2 replicon pRR10 (Roberts et al. (1990) *J Bacteriol.* 172(11): 6204–6216), yielding vector pAM132. Expression plasmid pAM97 was digested to completion using *AscI* and *SacI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 9.4 kb DNA fragment was gel extracted, and the isolated DNA fragment was ligated into the *MluI* *SacI* restriction enzyme site of pAM132, yielding expression plasmid pAM128.

Example 3

- [00230] This example describes methods for making expression vectors that encode enzymes including enzymes of the MEV pathway from *Enterococcus faecalis* organized in operons.
- [00231] Plasmid pAM16 was generated by inserting the coding sequence of the *mvaE* gene of *Enterococcus faecalis* (GenBank accession number AF290092 REGION: 1479..3890) (encodes an acetyl-CoA acetyltransferase/HMG-CoA reductase (HMGR)) into the pBlueScripII-KS(+) vector. The coding sequence of the *mvaE* gene was PCR amplified from *Enterococcus faecalis* genomic DNA (ATCC 700802) using 5' phosphorylated primers 4-40 mvaEF BamHI (SEQ ID NO: 11) and 4-40 mvaERHindIII (SEQ ID NO: 12). (Note that primer 4-40 mvaEF BamHI changes the start codon of the *mvaE* gene from TTG to ATG in the amplified PCR product.) The resulting PCR product was ligated into the *SmaI* restriction enzyme site of pBlueScripII-KS(+) (Stratagene, La Jolla, CA), yielding expression plasmid pAM16.
- [00232] Plasmid pAM18 was generated by inserting the coding sequence of the *mvaS* gene of *Enterococcus faecalis* (GenBank accession number AF290092 REGION: 142..1293) (encodes a HMG-CoA synthase (HMGS)) into the pBlueScripII-KS(+) vector. The coding sequence of the *mvaS* gene was PCR amplified from *Enterococcus faecalis* genomic DNA (ATCC 700802) using 5' phosphorylated primers 4-40 mvaSF BglII (SEQ ID NO: 13) and 4-39 mvaSR BamHI (SEQ ID NO: 14), and the PCR product was ligated into the *SmaI* restriction enzyme site of pBlueScripII-KS(+) (Stratagene, La Jolla, CA), yielding expression plasmid pAM18.
- [00233] Expression plasmid pAM22 was generated by inserting the coding sequence of the *mvaE* gene of expression plasmid pAM16 into the pZE21-P_{L-lacO1} vector. Vector pZE21-P_{L-lacO1} is a derivative of vector pZE21-MCS-1 in which the tet promoter was replaced with the P_{L-lacO1} promoter (Lutz and Bujard (1997) *Nucl Acids Res.* 25:1203-1210). Expression plasmid pAM16 was digested to completion using *BamHI* and *HindIII* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 2.4 kb DNA fragment containing the *mvaE* coding sequence was gel extracted, and the isolated DNA

fragment was inserted into the *Bam*HI *Hind*III restriction enzyme site of pZE21-P_{L-lacO1}, yielding expression plasmid pAM22.

- [00234] Expression plasmid pAM33 was generated by inserting the coding sequence of the *mvaS* gene of expression plasmid pAM18 into expression plasmid pAM22. Expression plasmid pAM18 was digested to completion using *Bgl*II and *Bam*HI restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 1.2 kb DNA fragment containing the coding sequence of the *mvaS* gene was gel extracted, and the isolated DNA fragment was inserted into the *Bam*HI site of expression plasmid pAM22, yielding expression plasmid pAM33.
- [00235] Expression plasmid pAM34 was generated by inserting the *mvaS-mvaE* operon of expression plasmid pAM33 into vector pAM29. The *mvaS-mvaE* operon was isolated by partially digesting pAM33 using *Eco*RI restriction enzyme, digesting the resulting linear DNA fragment using *Mlu*I restriction enzyme, resolving the reaction mixture by gel electrophoresis, and gel extracting the approximately 3.6 kb DNA fragment. The vector backbone of pAM29 was obtained by digesting to completion expression vector pAM25 using *Mlu*I and *Eco*RI restriction enzymes, resolving the reaction mixture by gel electrophoresis, and gel extracting the approximately 2.1 kb DNA fragment. The two isolated DNA fragments were ligated, yielding expression plasmid pAM34.

Example 4

- [00236] This example describes methods for making expression plasmids that encode enzymes, including enzymes of the DXP pathway from *Escherichia coli* organized in operons.
- [00237] Expression plasmid pAM408 was generated by inserting genes encoding enzymes of the “top” DXP pathway into the pAM29 vector. Enzymes of the “top” DXP pathway include 1-deoxy-D-xylulose-5-phosphate synthase (encoded by the *dxs* gene of *Escherichia coli*), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (encoded by the *dxr* gene of *Escherichia coli*), 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (encoded by the *ispD* gene of *Escherichia coli*), and 4-diphosphocytidyl-2C-methyl-D-erythritol synthase (encoded by the *ispE* gene of *Escherichia coli*), which together transform pyruvate and D-glyceraldehyde-3-phosphate to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate. DNA fragments comprising nucleotide sequences that encode enzymes of the “top” DXP pathway were generated by PCR amplifying the coding sequences of the *dxs* (GenBank accession number U00096 REGION: 437539..439401), *dxr* (GenBank accession number U00096 REGION: 193521..194717), *ispD* (GenBank accession number U00096 REGION: 2869803..2870512), and *ispE* (GenBank accession number U00096 REGION 1261249..1262100) genes from *Escherichia coli* strain DH1 (ATCC #33849) with added optimal Shine Dalgarno sequences and 5' and 3' restriction enzyme sites using the PCR primers shown in SEQ ID NOS: 15-18. The PCR products were resolved by gel electrophoresis, gel extracted using a Qiagen (Valencia, CA) gel purification kit, digested to completion using appropriate restriction enzymes (*Xho*I and *Kpn*I for the PCR product comprising the *dxs* gene; *Kpn*I and *Apa*I for the PCR product comprising the *dxr* gene; *Apa*I and *Nde*I for the PCR product comprising the *ispD* gene; *Nde*I and *Mlu*I for the PCR product comprising the *ispE* gene), and purified using a Qiagen (Valencia, CA) PCR purification kit. Roughly equimolar amounts of each PCR product were then added to a ligation reaction to assemble the individual genes into an operon. From this ligation reaction, 1 µl of reaction mixture was used to PCR amplify 2 separate gene cassettes, namely the *dxs-dxr* and the *ispD-ispE* gene cassettes. The *dxs-dxr* gene cassette was PCR amplified using primers 67-1A-C (SEQ ID NO: 15) and 67-1D-C (SEQ ID NO: 18), and the *ispD-ispE* gene cassette was PCR amplified using primers 67-1E-C (SEQ ID NO: 19) and 67-1H-C (SEQ

ID NO: 22). The two PCR products were resolved by gel electrophoresis, and gel extracted. The PCR product comprising the *dxs-dxr* gene cassette was digested to completion using *XhoI* and *ApaI* restriction enzymes, and the PCR product comprising the *ispD-ispE* gene cassette was digested to completion using *ApaI* and *MluI* restriction enzymes, and the two PCR products were purified. Vector pAM29 was digested to completion using *Sall* and *MluI* restriction enzymes, and the two digested PCR products containing the “top” DXP pathway operon were ligated into the *Sall MluI* restriction enzyme site of the pAM29 vector, yielding expression plasmid pAM408 (see Figure 4 for a plasmid map).

[00238] Expression plasmid pAM409 was generated by inserting genes encoding enzymes of the “bottom” DXP pathway into the pAM369 vector. Enzymes of the “bottom” DXP pathway include 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (encoded by the *ispF* gene of *Escherichia coli*), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (encoded by the *ispG* gene of *Escherichia coli*), and isopentenyl/dimethylallyl diphosphate synthase (encoded by the *ispH* gene of *Escherichia coli*), which together transform 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate to IPP and DMAPP. IPP is also converted to DMAPP through the activity of isopentenyl diphosphate isomerase (encoded by the *idi* gene of *Escherichia coli*). DMAPP can be further converted to FPP through the activity of farnesyl diphosphate synthase (encoded by the *ispA* gene of *Escherichia coli*). An operon encoding enzymes of the “bottom” DXP pathway as well as an isopentenyl diphosphate isomerase and a farnesyl diphosphate synthase was generated by PCR amplifying the *ispF* (GenBank accession number U00096 REGION: 2869323..2869802), *ispG* (GenBank accession number U00096 REGION: 2638708..2639826), *ispH* (GenBank accession number U00096 REGION: 26277..27227), *idi* (GenBank accession number AF119715), and *ispA* (GenBank accession number D00694 REGION: 484..1383) genes from *Escherichia coli* strain DH1 (ATCC #33849) with added optimal Shine Dalgarno sequences and 5' and 3' restriction enzyme sites using the appropriate PCR primers. The PCR products were resolved by gel electrophoresis, gel extracted, digested with the appropriate restriction enzymes (*BamHI* and *ApaI* for the PCR product comprising the *ispF* gene; *KpnI* and *ApaI* for the PCR product comprising the *ispG* gene; *Sall* and *KpnI* for the PCR product comprising the *ispH* gene; *Sall* and *HindIII* for the PCR product comprising the *idi* gene; *HindIII* and *NcoI* for the PCR product comprising the *ispA* gene), and purified. Roughly equimolar amounts of each PCR product were then added to a ligation reaction to assemble the individual genes into an operon. From this ligation reaction, 1 µl of reaction mixture was used to PCR amplify 2 separate gene cassettes, namely the *ispF-ispG* and the *ispH-idi-ispA* gene cassettes. The *ispF-ispG* gene cassette was PCR amplified using primers 67-2A-C (SEQ ID NO: 23) and 67-2D-C (SEQ ID NO: 26), and the *ispH-idi-ispA* gene cassette was PCR amplified using primers 67-2E-C (SEQ ID NO: 27) and 67-2J-C (SEQ ID NO: 32). The two PCR products were resolved by gel electrophoresis, and gel extracted. The PCR product comprising the *ispF-ispG* gene cassette was digested to completion using *BamHI* and *KpnI* restriction enzymes, and the PCR product comprising the *ispH-idi-ispA* gene cassette was digested to completion using *KpnI* and *NcoI* restriction enzymes, and the two PCR products were purified. Vector pAM369 was created by assembling the p15A origin of replication from pAM29 and beta-lactamase gene for ampicillin resistance from pZE12-luc (Lutz and Bujard (1997) *Nucl Acids Res.* 25:1203-1210) with an oligonucleotide-generated lacUV5 promoter. Vector pAM369 was digested to completion using *BamHI* and *NcoI* restriction enzymes, and the 2 isolated PCR products containing the “bottom” DXP pathway operon were ligated into the *BamHI NcoI* restriction enzyme site of the pAM369 vector, yielding expression plasmid pAM409.

[00239] Expression plasmid pAM424, a derivative of expression plasmid pAM409 containing the broad-host range RK2 origin of replication, was generated by transferring the *lacUV5* promoter and the *ispFGH-idi-ispA* operon of pAM409 to the pAM257 vector. Vector pAM257 was generated as follows: the RK2 *par* locus was PCR-amplified from RK2 plasmid DNA (Meyer et al. (1975) *Science* 190:1226-1228) using primers 9-156A (SEQ ID NO: 33) and 9-156B (SEQ ID NO: 34), the 2.6 kb PCR product was digested to completion using *AatII* and *XhoI* restriction enzymes, and the DNA fragment was ligated into a plasmid containing the p15 origin of replication and the chloramphenicol resistance gene from vector pZA31-luc (Lutz and Bujard (1997) *Nucl Acids Res.* 25:1203-1210), yielding plasmid pAM37-*par*; pAM37-*par* was digested to completion using restriction enzymes *SacI* and *HindIII*, the reaction mixture was resolved by gel electrophoresis, the DNA fragment comprising the RK2 *par* locus and the chloramphenicol resistance gene was gel extracted, and the isolated DNA fragment was ligated into the *SacI HindIII* site of the mini-RK2 replicon pRR10 (Roberts et al. (1990) *J Bacteriol.* 172:6204-6216), yielding vector pAM133; pAM133 was digested to completion using *BglIII* and *HindIII* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 6.4 kb DNA fragment lacking the ampicillin resistance gene and *oriT* conjugative origin was gel extracted, and the isolated DNA fragment was ligated with a synthetically generated DNA fragment comprising a multiple cloning site that contained *PciI* and *XhoI* restriction enzyme sites, yielding vector pAM257. Expression plasmid pAM409 was digested to completion using *XhoI* and *PciI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, and the approximately 4.4 kb DNA fragment was gel extracted. Vector pAM257 was digested to completion using restriction enzymes *XhoI* and *PciI*, and the isolated DNA fragment containing the *lacUV5* promoter and *ispFGH-idi-ispA* operon was ligated into the *XhoI PciI* restriction enzyme site of the pAM257 vector, yielding expression plasmid pAM424 (see Figure 5 for a plasmid map).

Example 5

[00240] This example describes methods for making expression plasmids that encode enzymes that convert FPP or GPP.

[00241] Expression plasmid pTrc99A-ADS was generated by inserting a nucleotide sequence encoding an amorpha-4,11-diene synthase ("ADS") into vector pTrc99A. The amorpha-4,11-diene synthase sequence was generated synthetically, so that upon translation the amino acid sequence would be identical to that described by Merke et al. (2000) *Ach. Biochem. Biophys.* 381:173-180, so that the nucleotide sequence encoding the amorpha-4,11-diene synthase was optimized for expression in *Escherichia coli*, and so that the nucleotide sequence was flanked by a 5' *NcoI* and a 3' *XmaI* restriction enzyme site (see U.S. Patent Number 7,192,751). The nucleotide sequence was digested to completion using *NcoI* and *XmaI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 1.6 kb DNA fragment was gel-extracted, and the isolated DNA fragment was inserted into the *NcoI XmaI* restriction enzyme site of the pTrc99A vector (Amman et al. (1985) *Gene* 40:183-190), yielding expression plasmid pTrc99A-ADS (see Figure 6 for a plasmid map).

[00242] Expression plasmid pAM113 is a chloramphenicol-resistant derivative of pTrc99A-ADS. It was generated by PCR amplifying the chloramphenicol resistance gene from vector pZA31-luc (Lutz and Bujard (1997) *Nucl Acids Res.* 25:1203-1210) using 5'-phosphorylated primers 19-137 cml-pAM37-AS (SEQ ID NO: 35) and 19-137 cml-pAM37-S (SEQ ID NO: 36), and inserting the 920 bp PCR product into the *FspI* restriction enzyme site of expression plasmid pTrc99A-ADS, yielding expression plasmid pAM113.

- [00243] Expression plasmid pC9 was generated by inserting a genomic DNA fragment of *Bacillus subtilis* 6051 comprising the coding sequence of the *nudF* gene and upstream genomic sequences (GenBank accession number Z99116 REGION: 49364..48548) into vector pTrc99A (Amann et al. (1988) *Gene* 69:301-315). Expression plasmid pNudF-H was generated by inserting the coding sequence of the *Bacillus subtilis* 6051 *nudF* gene (GenBank accession number Z99116 REGION: 49105..48548) into vector pTrc99A. Expression plasmid pyhfR was generated by inserting the coding sequence of the *Bacillus subtilis* 6051 *yhfr* gene (GenBank accession number Z99109 REGION: 97583..97002) into vector pTrc99A.
- [00244] Expression plasmid pAM373 was generated by inserting a nucleotide sequence encoding the β -farnesene synthase ("FSB") of *Artemisia annua* (GenBank accession number AY835398), codon-optimized for expression in *Escherichia coli*, into the pTrc99A vector. The nucleotide sequence encoding the β -farnesene synthase was generated synthetically, and was amplified by PCR from its DNA synthesis construct using the appropriate primers. To create a leader *NcoI* restriction enzyme site in the PCR product comprising the β -farnesene synthase coding sequence, the codon encoding the second amino acid in the original polypeptide sequence (TCG coding for serine) was replaced by a codon encoding aspartic acid (GAC) in the 5' PCR primer (SEQ ID NO: 37). The resulting PCR product was partially digested using *NcoI* restriction enzyme, and digested to completion using *SacI* restriction enzyme, the reaction mixture was resolved by gel electrophoresis, the approximately 1.7 kb DNA fragment comprising the β -farnesene synthase coding sequence was gel extracted, and the isolated DNA fragment was ligated into the *NcoI SacI* restriction enzyme site of the pTrc99A vector, yielding expression plasmid pAM373 (see Figure 6 for a plasmid map).
- [00245] Expression plasmids pTrc99A-FSA, pTrc99A-GTS, pTrc99A-PS, pTrc99A-TS were generated by inserting a DNA fragment comprising a nucleotide sequence encoding an α -farnesene synthase ("FSA"), a γ -terpinene synthase ("GTS"), an α -pinene synthase ("APS"), or a terpinolene synthase ("TS") into the pTrc99A vector. The DNA fragment insert was generated synthetically, using as a template for example the coding sequence of the α -farnesene synthase gene of *Picea abies* (GenBank accession number AY473627, REGION: 24..1766), the coding sequence of the β -farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398), the coding sequence of the γ -terpinene synthase gene of *Citrus limon* (GenBank accession number AF514286 REGION: 30..1832), the coding sequence of the α -pinene synthase gene of *Abies grandis* (GenBank accession number U87909, REGION: 6..1892) or of *Pinus taeda* (GenBank accession number AF543530 REGION: 1..1887), or the coding sequence of the terpinolene synthase gene of *Ocimum basilicum* (GenBank accession number AY693650) or of *Pseudotsuga menziesii* (GenBank accession number AY906866 REGION: 10..1887) or of *Abies grandis* (GenBank accession number AF139206), all nucleotide sequences being codon-optimized for expression in *Escherichia coli*. The DNA fragments for FSA was amplified by PCR from its DNA synthesis construct using the primer sequences SEQ ID NO: 39 and SEQ ID NO: 40. The resulting PCR product was digested to completion using *NcoI* and *SacI* restriction enzymes, the reaction mixture was resolved by gel electrophoresis, the approximately 1.7 kb DNA fragment comprising the α -farnesene synthase coding sequence was gel extracted, and the isolated DNA fragment was ligated into the *NcoI SacI* restriction enzyme site of the pTrc99A vector, yielding expression plasmid pTrc99A-FSA (see figure 6 for a plasmid map). The DNA fragments for GTS, APS, and TS were designed to be flanked by a leader *XmaI* restriction enzyme site and a terminal *XbaI* restriction enzyme site, and were cloned into compatible restriction enzyme sites of a cloning vector such as a standard pUC or pACYC origin vector, from which they could be liberated again

by digesting to completion the DNA synthesis construct using *XbaI* and *XmaI* restriction enzymes, resolving the reaction mixture by gel electrophoresis, and gel extracting the 1.7 to 1.9 terpene synthase encoding DNA fragment. The isolated DNA fragments were ligated into the *XmaI XbaI* restriction enzyme site of vector pTrc99A (Amman *et al.*, *Gene* 40:183-190 (1985)), yielding plasmids pTrc99A-GTS, pTrc99A-APS, or pTrc99A-TS (see Figure 6 for plasmid maps).

- [00246] Expression plasmids pRS425-FSA and pRS425-FSB were generated by inserting a nucleotide sequence encoding an α -farnesene synthase ("FSA") or a β -farnesene synthase ("FSB"), respectively, into the pRS425-Gal1 vector (Mumberg *et al.* (1994) *Nucl. Acids. Res.* 22(25): 5767-5768). The nucleotide sequence inserts were generated synthetically, using as a template for example the coding sequence of the α -farnesene synthase gene of *Picea abies* (GenBank accession number AY473627, REGION: 24..1766) or of the β -farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398), codon-optimized for expression in *Saccharomyces cerevisiae*. The synthetically generated nucleotide sequence was flanked by a 5' *BamHI* site and a 3' *XhoI* site, and could thus be cloned into compatible restriction enzyme sites of a cloning vector such as a standard pUC or pACYC origin vector. The synthetically generated nucleotide sequence was isolated by digesting to completion the DNA synthesis construct using *BamHI* and *XhoI* restriction enzymes. The reaction mixture was resolved by gel electrophoresis, the approximately 1.7 kb DNA fragment comprising the α -farnesene synthase or β -farnesene synthase coding sequence was gel extracted, and the isolated DNA fragment was ligated into the *BamHI XhoI* restriction enzyme site of the pRS425-Gal1 vector, yielding expression plasmid pRS425-FSA or pRS425-FSB, respectively.
- [00247] Expression plasmids pTrc99A-LLS, pTrc99A-LMS, pTrc99A-BPS, pTrc99A-PHS, pTrc99A-CS, and pTrc99A-SS are generated by inserting a nucleotide sequence encoding a linalool synthase ("LLS"), limonene synthase ("LMS"), β -pinene synthase ("BPS"), β -phellandrene ("PHS"), carene synthase ("CS"), or sabinene synthase ("SS") into the pTrc99A vector. The nucleotide sequence inserts are generated synthetically, using as a template for example the coding sequence of the linalool synthase gene of *Artemisia annua* (GenBank accession number AF154124, REGION: 13..1764), the coding sequence of the limonene synthase gene of *Abies grandis* (GenBank accession number AF006193 REGION: 73..1986), the coding sequence of the β -pinene synthase of *Artemisia annua* (GenBank accession number AF276072 REGION: 1..1749), the coding sequence of the β -phellandrene synthase gene of *Abies grandis* (GenBank accession number AF139205 REGION: 34..1926), the coding sequence of the carene synthase gene of *Salvia stenophylla* (GenBank accession number AF527416 REGION: 78..1871), or the coding sequence of the sabinene synthase gene of *Salvia officinalis* (GenBank accession number AF051901 REGION: 26..1798). The nucleotide sequences encoding the β -pinene, sabinene, and β -phellandrene synthases are flanked by a leader *XmaI* restriction enzyme site and a terminal *XbaI* restriction enzyme site, the nucleotide sequences encoding the linalool and carene synthases are flanked by a leader *NcoI* restriction enzyme site and a terminal *XmaI* restriction enzyme site, and the nucleotide sequence encoding the limonene synthase is flanked by a leader *NcoI* restriction enzyme site and a terminal *PstI* restriction enzyme site. The DNA synthesis constructs are digested to completing using *XmaI* and *XbaI* (for the β -pinene, sabinene, and β -phellandrene synthase constructs), *NcoI* and *XmaI* restriction enzymes (for the linalool and carene synthase constructs), or *XbaI* and *PstI* restriction enzymes (for the limonene synthase construct). The reaction mixtures are resolved by gel electrophoresis, the approximately 1.7 to 1.9 kb DNA fragments are gel extracted, and the isolated DNA fragments are ligated into the *XmaI XbaI* restriction enzyme site (for the β -

pinene, sabinine, and β -phellandrene synthase inserts), the *NcoI* *XmaI* restriction enzyme site (for the linalool and carene synthase inserts), or the *XbaI* *PstI* restriction enzyme site (for the limonene synthase insert) of the pTrc99A vector, yielding expression plasmids pTrc99A-LLS, pTrc99A-LMS, pTrc99A-BPS, pTrc99A-PHS, pTrc99A-CS, and pTrc99A-SS (see Figure 6 for plasmid maps).

Example 6

[00248] This example describes the generation of *Escherichia coli* host strains useful in the invention.

[00249] As detailed in Table 1, the host strains were created by transforming chemically competent *Escherichia coli* parent cells with one or more expression plasmids of Example 1 through 5.

Table 1. *E. coli* host strains

Host Strain	E.coli Parent Strain	Expression Plasmids	Antibiotic Selection
B32	DH1	pMevT	100 ug/mL carbenicillin
B292	B	pMBIS	5 ug/mL tetracycline
B210	DP	pTrc99A-ADS	34 ug/mL chloramphenicol
B153	DH1	pAM97	100 ug/mL carbenicillin
B282	DP	pTrc99A-ADS	34 ug/mL chloramphenicol
B255	DH1	pAM128	100 ug/mL carbenicillin
B256	DP	pAM113	34 ug/mL chloramphenicol
B86	DH1	pAM52 pMBIS pTrc99A-ADS	50 ug/mL kanamycin 100 ug/mL carbenicillin 5 ug/mL tetracycline
B61	DH1	pAM25 pBBR1MCS-3 pTrc99A	
B62		pAM34 pBBR1MCS-3 pTrc99A	
B003	DH10B	pTrc99A-ADS	100 ug/mL carbenicillin
B617		pAM408 pTrc99A-ADS	100 ug/mL carbenicillin 50 ug/mL kanamycin
B618		pAM424 pTrc99A-ADS	100 ug/mL carbenicillin 35 ug/mL chloramphenicol
B619		pAM408 pAM424 pTrc99A-ADS	100 ug/mL carbenicillin 50 ug/mL kanamycin 35 ug/mL chloramphenicol
B650	DH10B	pAM373	100 ug/mL carbenicillin
B651		pAM408 pAM373	100 ug/mL carbenicillin 50 ug/mL kanamycin
B652		pAM424 pAM373	100 ug/mL carbenicillin 35 ug/mL chloramphenicol
B653		pAM408 pAM424 pAM373	100 ug/mL carbenicillin 50 ug/mL kanamycin 35 ug/mL chloramphenicol
B286	DH1	pAM97-MevB pC9	100 ug/mL carbenicillin 34 ug/mL chloramphenicol.
B287		pAM97-MevB pnudF-H	
B288		pAM97-MevB pyhfR	
B291		pAM97-MBI pyhfR	
B592	DH1	pMevT pMBIS pTrc99A-FSA	100 ug/mL carbenicillin 34 ug/mL chloramphenicol
B552		pMevT	5 ug/mL tetracycline

Host Strain	E.coli Parent Strain	Expression Plasmids	Antibiotic Selection
		pMBIS pAM373	
Example 21 host cell (production of GTS, APS, TS)		pMevT pMBIS-gpps pTrc99A-GTS or - APS or -TS	
Example 21 host cell (production of LLS, LMS, BPS, PHS, CS, SS)		pMevT pMBIS-gpps pTrc99A-LLS or - LMS or -BPS or - PHS or -CS or -SS	100 ug/mL carbenicillin 34 ug/mL chloramphenicol 5 ug/mL tetracycline

[00250] Host cell transformants were selected on Luria Bertoni (LB) agar containing antibiotics as detailed in Table

1. Single colonies were transferred from LB agar to culture tubes containing 5 mL of LB liquid medium and antibiotics. B003, B617, B618, B619, B650, B651, B652, and B653 host cell transformants were incubated at 30°C on a rotary shaker at 250 rpm for 30 hours. All other host cell transformants were incubated at 37°C on a rotary shaker at 250 rpm until growth reached stationary phase. The cells were adapted to minimal media by passaging them through 4 to 5 successive rounds of M9-MOPS media containing 0.8% glucose and antibiotics (see Table 2 for the composition of the M9-MOPS medium). The cells were stored at -80°C in cryo-vials in 1 mL stock aliquots made up of 400 uL sterile 50% glycerol and 600 uL liquid culture.

Table 2 – Composition of M9-MOPS Culture Medium

Component	Quantity (per L)
Na ₂ HPO ₄ 7H ₂ O	12.8 g
KH ₂ PO ₄	3 g
NaCl	0.5 g
NH ₄ Cl	1 g
MgSO ₄	2 mmol
CaCl ₂	0.1 mmol
Thiamine	0.1 ug
MOPS buffer pH 7.4	100 mmol
(NH ₃) ₆ Mo ₇ O ₂₄ 4H ₂ O	3.7 ug
H ₃ BO ₄	25 ug
CoCl ₂	7.1 ug
CuSO ₄	2.4 ug
MnCl ₂	16 ug
ZnSO ₄	2.9 ug
FeSO ₄	0.28 mg

Example 7

[00251] This example demonstrates expression plasmid stability in the absence of antibiotics in an *Escherichia coli* host strain that harbors an expression plasmid comprising the RK2 plasmid replication, segregation, and maintenance system.

- [00252] A seed culture of host strain B255 was established by adding a stock aliquot of the strain to a 125 mL flask containing 40 mL M9-MOPS, 2% glucose, 0.5% yeast extract, and antibiotics as detailed in Table 1, and by growing the culture overnight.
- [00253] The seed culture was used to inoculate at an initial OD₆₀₀ of approximately 0.05, two 250 mL flasks each containing 40 mL M9-MOPS medium, 2% glucose, and 0.5% yeast extract. Culture #1 also contained 100 ug/mL carbenicillin and 34 ug/mL chloramphenicol. Culture #2 did not receive any antibiotics. Both cultures were incubated at 37°C on a rotary shaker at 250 rpm until they reached an OD₆₀₀ of approximately 0.2, at which point the production of amorpho-4,11-diene in the host cells was induced by adding 40 uL of 1M IPTG to the culture medium. At the time of induction, the cultures were overlain with 8 mL of an organic overlay to capture the amorpho-4,11-diene. Samples were taken periodically for a total of 72 hours. Production of amorpho-4,11-diene by the host strain in the 2 cultures was confirmed by GC/MS as described in Example 10.
- [00254] To assess plasmid stability in the two cell cultures, a sample of each culture was removed at 72 hours and streaked onto a LB agar plate (no antibiotics). After overnight incubation at 37°C, 50 individual colonies derived from each culture were replica-plated onto a LB agar-plus-antibiotics (34 ug/mL chloramphenicol, 100 ug/mL carbenicillin) plate and a LB agar-minus-antibiotics (no antibiotic) plate. After another overnight incubation at 37°C, the LB agar-plus-antibiotics and the LB agar-minus-antibiotics plate were each found to contain approximately 50 colonies, indicating that plasmid retention both in the presence and in the absence of antibiotics in the culture medium had been approximately 100%.

Example 8

- [00255] This example demonstrates increased specific activity and stability of the *Enterococcus faecalis* HMGR compared to the *Saccharomyces cerevisiae* tHMGR in an *Escherichia coli* host strain.
- [00256] Seed cultures of host strains B61 and B62 were established by adding a stock aliquot of each strain to 125 mL flasks containing 20 mL M9-MOPS medium, 0.8% % glucose, and antibiotics as detailed in Table 5, and by growing the cultures to saturation. The seed cultures were diluted 1:100 into 140 mL of fresh medium in a 500 mL flask, and grown again to an OD₅₅₀ of approximately 0.1, at which point production of amorpho-4,11-diene was induced by adding 140 uL 1 M IPTG to each culture. At 4, 12, 20, 28, 36, and 49 hours post-induction, samples were removed from each culture, and cells were pelleted by centrifugation. The cell pellets were snap frozen on dry ice, and then stored at -80°C.
- [00257] To conduct enzyme assays, cell pellets were thawed on ice, and then lysed using Bugbuster (Novagen, Madison, WI) containing protease inhibitor mix #3 (Calbiochem, San Diego, CA), benzonase (20 uL oer5 mL bugbuster; Novagen, Madison, WI), and lysozyme (30 ug/mL). Enzyme activity of the *Saccharomyces cerevisiae* tHMGR was assayed in 50 mM Tris HCl (pH7.5), 0.2 mM NADPH (Sigma, St. Louis, MO), and 0.3 mM DL-3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) sodium salt (Sigma, St. Louis, MO). The assay was started by adding cell lysate, and the disappearance of NADPH was monitored by absorbance at 340nm. To account for non-specific disappearance of NADPH, results obtained in a control assay lacking HMG-CoA were subtracted from results obtained in test samples. Enzyme activity of the *Enterococcus faecalis* HMGR was measured similarly except that the assay buffer contained 100 mM potassium phosphate buffer (pH6.5), 0.4 mM NADPH, 1.0 mM EDTA, and 100 mM KCl.
- [00258] Protein assays were done by the method of Bradford ((1976) *Anal Biochem.* 72:248-254). Specific activities were calculated as Δnmol NADPH/min/mg protein.

[00259] As shown in Figure 8, the *Enterococcus faecalis* HMGR exhibited higher specific activity and increased stability compared to the *Saccharomyces cerevisiae* tHMGR.

Example 9

[00260] This example describes the calibration of OD₆₀₀ with dry cell weight ("DCW").

[00261] To obtain the relationship between DCW and OD₆₀₀, a representative strain, B32, was grown in high cell density processes similar to those described in Examples 10-14. Samples were taken throughout the runs, and the OD₆₀₀ and DCW were measured for each sample. To determine the DCW, the cells were pelleted and the supernatant discarded. The cell pellet was washed once with water, and was then dried in an oven at 80°C for at least 3 days. The tubes containing cell pellets were weighed, the weight of the tube was subtracted from the measured weights, and the remaining weight was divided by the initial volume of each sample (0.0015 L) to obtain the DCW.

[00262] Figure 9 shows the relationship between DCW and OD₆₀₀ measured in these experiments.

Example 10

[00263] This example demonstrates increased production of amorpho-4,11-diene in *Escherichia coli* host strains expressing the *Staphylococcus aureus* HMGR and HMGS compared to host strains expressing the *Saccharomyces cerevisiae* tHMGR and HMGS.

[00264] Seed cultures of host strains B32, B153, B210, B282, B292, B86, B255, and B256 were established by adding a stock aliquot of each strain to separate 125 mL flasks containing 25 mL M9-MOPS medium, 0.8% glucose, and antibiotics as detailed in Table 1, and by growing the cultures overnight.

[00265] The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05 separate 250 mL flasks containing 40 mL M9-MOPS medium, 2% glucose, and antibiotics. The cultures were incubated at 30°C on a rotary shaker at 250 rpm until they reached an OD₆₀₀ of approximately 0.2, at which point the production of amorpho-4,11-diene in the host cells was induced by adding 40 uL of 1M IPTG to the culture medium. The cultures were overlain with 8mL of an organic overlay (e.g., dodecane, methyl oleate or isopropyl myristate). Samples of the organic overlay layer and the broth were taken once a day for 72 hours. Broth samples were used to measure the OD₆₀₀. Amorpho-4,11-diene concentration was measured by transferring 5 uL of the organic overlay layer to a clean glass vial containing 500 uL ethyl acetate spiked with beta- or trans-caryophyllene as an internal standard.

[00266] The organic overlay/ethyl acetate samples were analyzed on a Hewlett-Packard 6890 gas chromatograph/mass spectrometer (GC/MS) by scanning only for two ions, the molecular ion (204 m/z) and the 189 m/z ion, as described in Martin et al. (2001) *Biotechnol. Bioeng.* 75:497-503. To expedite run times, the temperature program and column matrix was modified to achieve optimal peak resolution and the shortest overall runtime. A 1 uL sample was separated on the GC using a DB-XLB column (available from Agilent Technologies, Inc., Palo Alto, CA) and helium carrier gas. The temperature program for the analysis was as follows: 100°C for 0.75 minutes, increasing temperature at 60°C/minute to a temperature of 300°C, and a hold at 300°C for 0.5 minutes. The resolved samples were analyzed by a Hewlett-Packard model 5973 mass-selective detector that monitored ions 189 and 204 m/z. Previous mass spectra demonstrated that the amorpho-4,11-diene synthase product was amorpho-4,11-diene, and that amorpho-4,11-diene had a retention time of 3.7 minutes using this GC protocol. Beta- or trans-caryophyllene was used as an internal standard for quantitation. Amorpho-4,11-diene titer was calculated using the ratio of internal standard to amorpho-4,11-diene peak areas based upon a quantitative

calibration curve of purified amorpho-4,11-diene (0.63-10 mg/L of KJF17-109-3) in caryophyllene-spiked ethyl acetate.

[00267] As shown in Figures 10A and 10B, strains B153 and B282, which expressed the *Staphylococcus aureus* HMGR and HMGS, produced elevated levels of amorpho-4,11-diene compared to strains B32, B210, B255, B256, and B292, which expressed the *Saccharomyces cerevisiae* HMGR and HMGS.

Example 11

[00268] This example demonstrates increased production of amorpho-4,11-diene by an *Escherichia coli* host strain grown at suboptimal temperature.

[00269] A seed culture of host strain B32 was established by adding 0.5 mL of a stock aliquot of the strain to a 250 mL flask containing 50 mL M9-MOPS medium and antibiotics as detailed in Table 1, and by growing the culture overnight at 37°C on a rotary shaker at 250 rpm.

[00270] The seed culture was used to inoculate at an initial OD₆₀₀ of approximately 0.05 four 250 mL flasks, each containing 40 mL fermentor batch medium (see Table 6 for medium composition), 100 mM MOPS buffer pH7.1, and antibiotics. The cultures were incubated on a rotary shaker at 250 rpm at either 30°C or 37°C until they reached an OD₆₀₀ of 0.18 to 0.22, at which point the production of amorpho-4,11-diene in the host cells was induced by adding 40 uL of 1M IPTG to the culture medium. At the time of induction, the cultures were overlain with 8mL of an organic overlay to capture the amorpho-4,11-diene. Samples were taken once a day, and analyzed as described in Example 10.

[00271] As shown in Figures 11A and 11B, fermentation at 30°C did not affect cell growth, but led to an approximately 50% increase in the specific production of amorpho-4,11-diene by the *Escherichia coli* host strain.

Example 12

[00272] This example demonstrates increased production of amorpho-4,11-diene by an *Escherichia coli* host strain grown under restricted carbon source conditions.

[00273] A seed culture of host strain B32 for fermentation runs 050608-1 and 050629-1 was established by adding 0.25 uL of a stock aliquot of the strain to a 250 mL flask containing 50 mL M9-MOPS medium and antibiotics as detailed in Table 1, and by incubating the culture at 37°C on a rotary shaker at 250 rpm until it reached an OD₆₀₀ of 1 to 2.

[00274] A seed culture of host strain B32 for fermentation run 060403-3 was established by adding a stock aliquot of the strain to a 250 mL flask containing 50 mL M9-MOPS medium and antibiotics as detailed in Table 1, and by incubating the culture overnight at 37°C on a rotary shaker at 250 rpm. The seed culture was used to inoculate at an initial OD₆₀₀ of approximately 1 a 250 mL flask containing 40 mL M9-MOPS medium and antibiotics, and the culture was again incubated at 37°C on a rotary shaker at 250 rpm until it reached an OD₆₀₀ of 3 to 5.

[00275] For all fermentation processes, the KH₂PO₄, K₂HPO₄ 3H₂O, EDTA, citric acid, and (NH₄)₂SO₄ were heat sterilized in the bioreactor (2L Applikon Bioconsole ADI 1025s with ADI 1010 controllers, Applikon Biotechnology, Foster City, CA). The remaining media components were filter sterilized as stock solutions and injected through the headplate. Table 3 shows the final media composition for fermentation runs 050608-1 and 050629-1. Table 4 shows the final media composition for fermentation run 060403-3. The starting volume for run 050608-1 was 0.8 L, the starting volume for 050629-1 was 1.2 L and the starting volume for 060403-3 was 1 L. All runs were inoculated by injecting 50 mL of the seed culture through the headplate.

TABLE 3 – Composition of Fermentation Medium of Fermentation
Runs 050608-1 and 050629-1

Component	Batch Medium (per L)	Feed Solution (per L)
Glucose	5 g	590-650 g
KH_2PO_4	4.2 g	-
$\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$	15.7 g	-
Citric acid	1.7 g	-
$(\text{NH}_4)_2\text{SO}_4$	2 g	-
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	1.2 g	12 g
EDTA	8.4 mg	13 g
$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$	0.25 mg	0.4 mg
$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$	1.5 mg	2.35 mg
$\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$	0.15 mg	0.25 mg
H_3BO_4	0.3 mg	0.5 mg
$\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$	0.25 mg	0.4 mg
$\text{Zn}(\text{CH}_3\text{COO})_2 \cdot 2\text{H}_2\text{O}$	1.3 mg	1.6 mg
Fe(III)citrate hydrate	10.0 mg	4.0 mg
Thiamine HCl	4.5 mg	-
Carbenicillin	100 ug	100 ug
Tetracycline	5 ug	5 ug
Chloramphenicol	34 ug	34 ug

Table 4 – Composition of Fermentation Medium of Fermentation Run 060403-3

Component	Batch medium (per L)	Feed solution (per L)
Glucose	15 g	650 g
KH ₂ PO ₄	4.2 g	-
K ₂ HPO ₄ 3H ₂ O	15.7 g	-
Citric acid	1.7 g	-
(NH ₄) ₂ SO ₄	2 g	-
MgSO ₄ 7H ₂ O	1.2 g	12 g
EDTA	8.4 mg	13 mg
CoCl ₂ 6H ₂ O	2.5 mg	4 mg
MnCl ₂ 4H ₂ O	15 mg	23.5 mg
CuCl ₂ 2H ₂ O	1.5 mg	2.5 mg
H ₃ BO ₄	3 mg	5 mg
Na ₂ MoO ₄ 2H ₂ O	2.5 mg	4 mg
Zn(CH ₃ COO) ₂ 2H ₂ O	13 mg	16 mg
Fe(III)citrate hydrate	100 mg	40 mg
Thiamine HCl	4.5 mg	-
Carbenicillin	100 ug	100 ug
Tetracycline	5 ug	5 ug
Chloramphenicol	34 ug	34 ug

[00276] For fermentation run 050608-1 (excess carbon), the feed was initiated at induction, and feed rates were adjusted manually to provide glucose in the concentrations shown in Figure 12C. For fermentation run 050629-1 (carbon-restricted), the feed was delivered to the fermentor according to the protocol shown in Table 5. For fermentation run 060403-3 (lowest carbon), the feed was started automatically when the initial glucose bolus (15 g) was exhausted and the dissolved oxygen spiked. Up to a maximum of 27.6 g/hr, the rate of the feed was calculated according to the following equation:

$$m_s(t) = S(t_0)\mu e^{\mu(t-t_0)}$$

$$\mu = 0.12$$

$$S(t_0) = 15g$$

wherein t_0 is the time at which the initial glucose was depleted. Upon reaching the maximum rate, the glucose feed was restricted to a rate of 9.5 g/hr, and held constant at this rate for the remainder of the run.

Table 5 – Feed Protocol for Fermentation Run 050629-1

Run Time (hours)	Glucose Feed Rate (g/hr)
0	0
7	0.37
10	0.74
12	1.11
14	1.48
16	2.22
18	2.96
20	3.69
22	4.80
24	5.91
31	7.39
33	5.54
47	3.69

- [00277] Runs 050608-1 and 050629-1 were carried out at 37°C. Airflow in the bioreactor was set at 1-2 L/min; pH was maintained at 7 using ammonium hydroxide and/or sodium hydroxide; initial agitation was 500-600 rpm; foam was controlled with antifoam B (Sigma-Aldich, St. Louis, MO); the dissolved oxygen levels were maintained above 30% using an agitation cascade. After 5-6 hours of cultivation, production of amorpho-4,11-diene by the host cells was induced by adding 0.8 mL of 1 M IPTG to run 050608-1 and 1.2 mL IPTG to run 050629-1. Upon induction, the culture temperature was reduced to 30°C.
- [00278] Run 060403-3 was carried out at 30°C. Airflow in the bioreactor was set at 1-2 L/min; pH was maintained at 7 using ammonia hydroxide. Dissolved oxygen was maintained above 30% by an agitation cascade and oxygen enrichment. At an OD₆₀₀ of approximately 28 (19 hours after inoculation), production of amorpho-4,11-diene by the host cells was induced by adding 1 mL 1 M IPTG.
- [00279] Amorpho-4,11-diene was captured and extracted according to two different protocols. For runs 050608-1 and 050629-1, volatile amorpho-4,11-diene present in the off-gas was captured by venting the off-gas through a gas-washer containing 200 mL heptanol. The heptanol was then diluted into ethyl acetate until the amorpho-4,11-diene concentration in the sample was between 0.63 mg/L and 20 mg/L. For run 060403-3, amorpho-4,11-diene was captured in the bioreactor by adding 200 mL of an organic overlay to the fermentor at the time of induction. Product concentration was measured by combining 25 uL broth plus organic overlay with 975 uL acetonitrile, shaking the sample at maximum speed on a Fisher Vortex Genie 2™ mixer (Scientific Industries, Inc., Bohemia, NY) for at least 3 minutes, removing cells from the sample by centrifugation, and diluting the acetonitrile solution into ethyl acetate until the amorpho-4,11-diene concentration in the sample was between 0.63 and 20 mg/L. The ethyl acetate samples were analyzed by GC/MS as described in Example 10.
- [00280] As shown in Figures 12A and 12B, fermentation run 050608-1 (excess carbon) resulted in low maximum cell densities and low production of amorpho-4,11-diene, respectively, correlating, at least in part, to the relatively rapid increase in acetate levels (Figure 12D). In comparison, fermentation run 050629-1 (carbon-restricted) resulted in increased production of amorpho-4,11-diene (Figure 12B), and delayed the onset of

acetate production. These results are consistent with the hypothesis that excess glucose feeds lead to rapid acetate production and early cell death.

[00281] Further glucose restriction as achieved by fermentation run 060403-3 (lowest carbon) resulted in low acetate production for over 100 hours (Figure 12D), and significantly higher maximum cell density and amorpha-4,11-diene production (Figures 12A and 12B).

Example 13

[00282] This example demonstrates increased amorpha-4,11-diene production by an *Escherichia coli* host strain grown under restricted carbon source conditions and at suboptimal temperature.

[00283] A seed culture of host strain B153 was established by adding a stock aliquot of the strain to a 250 mL flask containing 50 mL M9-MOPS medium and antibiotics as detailed in Table 1, and growing the culture at 37°C on a rotary shaker at 250 rpm to an OD₆₀₀ of 3.5 to 4.5.

[00284] 2 L bioreactors (Biocontroller ADI 1010 with Bioconsole ADI 1025, Applikon Biotechnology, Foster City, CA) were set up and run in the same way as described in Example 12 for run 060403-3, except that strain and induction time were varied.

[00285] Production of amorpha-4,11-diene in the host cells was induced by adding 1 mL of 1 M IPTG to the culture medium. In the fermentation run shown in Figure 13A, amorpha-4,11-diene synthesis was induced at an OD₆₀₀ of approximately 2, while the fermentor still contained excess glucose. In the fermentation run shown in Figure 13B, amorpha-4,11-diene synthesis was induced at an OD₆₀₀ of approximately 33, which was after the glucose-restricted feed had started.

[00286] Amorpha-4,11-diene was captured and extracted according to two different protocols. For the fermentation run shown in Figure 13A, volatile amorpha-4,11-diene present in the off-gas was captured by venting the off-gas through a gas-washer containing 200 mL heptanol. The heptanol was then diluted into ethyl acetate until the amorpha-4,11-diene concentration in the sample was between 0.63 and 20 mg/L. For the fermentation run shown in Figure 13B, amorpha-4,11-diene was captured by adding 200 mL of an organic overlay to the fermentor at the time of induction.

[00287] Amorpha-4,11-diene was extracted from the culture medium by combining 25 uL broth with 975 uL acetonitrile, shaking the sample at maximum speed on a Fisher Vortex Genie 2™ mixer (Scientific Industries, Inc., Bohemia, NY) for at least 3 minutes, removing cells from the sample by centrifugation, and diluting the acetonitrile solution into ethyl acetate until the amorpha-4,11-diene concentration in the sample was between 0.63 and 20 mg/L. The ethyl acetate samples were analyzed by GC/MS as described in Example 10. For the fermentation run shown in Figure 13A, the total amount of amorpha-4,11-dien was derived by combining the amounts present in the off-gas and in the culture medium, and dividing the total by the fermentor volume.

[00288] The fermentation shown in Figure 13A reached a maximal OD₆₀₀ of 93 and a maximal amorpha-4,11-diene concentration of 3.2 g/L. In contrast, the fermentation shown in Figure 13B reached a maximal OD₆₀₀ of 245 and a maximal amorpha-4,11-diene concentration of 15 g/L. A likely explanation for the differences in culture growth and amorpha-4,11-diene production levels observed in the two cultures is that in the fermentation run shown in Figure 13A amorpha-4,11-diene production was induced before the excess glucose was consumed, and that the unrestricted availability of glucose caused cell death by enabling the build-up of toxic levels of intermediates of the mevalonate pathway. In the fermentation run shown in Figure 13B, induction occurred after glucose delivery was restricted, which prevented the build-up of pathway intermediates, leading to higher cell density and amorpha-4,11-diene production levels.

Example 14

- [00289] This example demonstrates increased amorpho-4,11-diene production by an *Escherichia coli* host strain grown under restricted carbon and nitrogen source conditions and at suboptimal temperature.
- [00290] A seed culture of host strain B86 was established by adding a stock aliquot of the strain to a 250 mL flask containing 50 mL M9-MOPS medium and antibiotics as detailed in Table 1. The culture was grown overnight at 37°C on a rotary shaker at 250 rpm, sub-cultured the following morning into the same medium at an OD₆₀₀ of approximately 1, and grown again at 37°C and 250 rpm to an OD₆₀₀ of 3 to 5.
- [00291] Four 2 L bioreactors (Biocontroller ADI 1010 with Bioconsole ADI 1025, Applikon Biotechnology, Foster City, CA) were set up and run in the same way as described in Example 12 for run 060403-3, except that the nitrogen restricted runs did not contain ammonia sulfate in the feed.
- [00292] An exponential glucose feed with a 6 hour doubling time was initiated automatically when the initial glucose bolus (15 g) was exhausted and the dissolved oxygen spiked. Up to a maximum of 30.4 g/hr, the rate of the feed was calculated according to the following equation:

$$m_s(t) = S_0 \mu e^{\mu(t-t_0)}$$
$$\mu = 0.12 \text{ min}^{-1}$$
$$S_0 = 15 \text{ g}$$

wherein μ is the specific growth rate, and t_0 is the time at which the initial glucose bolus was depleted. Upon reaching the maximum rate, the glucose feed was reduced to a rate of 11.4 g/hr, and held constant at this rate for the remainder of the run. In fermentation runs 060710-4, 060724-5, and 060619-5 (carbon- and nitrogen-restricted), the glucose feed was further reduced when ammonia restriction lead to glucose accumulation in the medium.

- [00293] Fermentation was carried out at the reduced temperature of 30°C. Airflow in the bioreactor was set at 1 vvm; initial agitation was at 700 rpm; foam was controlled with antifoam B (Sigma-Aldich, St. Louis, MO); and dissolved oxygen tension was controlled at 40% using an agitation cascade (700-1,200 rpm) and oxygen enrichment. In fermentation run 060327-3 (carbon-restricted), the pH was maintained at 7 using 20% NH₄OH; in fermentation runs 060710-4, 060724-5, and 060619-5 (carbon- and nitrogen-restricted), pH was maintained at 7 initially using 20% NH₄OH, and starting at 72 hours using a 50/50 mixture of 2.5 N NaOH and 10 N NH₄OH, to further restrict the amount of ammonia going into the fermentor.
- [00294] Production of amorpho-4,11-diene in the host cells was induced at an OD₆₀₀ of approximately 30 by adding 1 mL of 1 M IPTG to the culture medium.
- [00295] Amorpho-4,11-diene was captured by overlaying the medium with 10% (v/v) of an organic overlay. Amorpho-4,11-diene was then extracted by combining 25 uL of broth with 975 uL methanol, shaking the sample at maximum speed on a Fisher Vortex Genie 2™ mixer (Scientific Industries, Inc., Bohemia, N.Y.) for at least 15 minutes, removing cells from the sample by centrifugation, and adding 10 uL of the methanol solution to 990 uL ethyl acetate containing 10 uL/L trans-caryophyllene.
- [00296] Samples were analyzed by GC/MS as described in Example 10.
- [00297] Figures 14A-E show data from fermentation run 060327-3 (carbon-restricted). The fermentation produced a maximum concentration of amorpho-4,11-diene of 16 g/L (Figure 14A). The maximum volumetric productivity of the host strain was more than 200 mg/L/hour (Figure 14B). The maximum specific productivity of the host strain was >2 mg/L/hour/OD₆₀₀ (Figure 14C). The concentration of ammonia in the culture medium was about 30 mM at the start of the fermentation run, rose to about 76 mM upon addition

of the feed solution during the exponential growth phase, and remained above 60 mM for the remainder of the run (Figure 14D). The maximum OD₆₀₀ reached was about 290 (Figure 14D), corresponding to 116 g DCW/L. The concentration of glucose in the culture medium dropped from 15 g/L to below 1 g/L in less than 20 hours, and remained low (Figure 14E). Acetate levels were low throughout the fermentation (Figure 14E).

[00298] Figures 15A-E show data from fermentation runs 060710-4, 060724-5, and 060619-5 (carbon- and nitrogen-restricted). The fermentations produced a maximum concentration of amorpho-4,11-diene from about 20 g/L to 30 g/L (Figure 15A). The maximum volumetric productivity of the host strain was more than 400 mg/L/hour in all three fermentation runs (Figure 15B), which is significantly higher than the maximum volumetric productivity obtained in the nitrogen unrestricted fermentation (Figure 14B). The maximum specific productivity of the host strain was >2 mg/L/hour/OD₆₀₀ for all runs, and remained high throughout the runs (Figure 15C). The concentration of ammonia in the culture medium was about 35 mM to 50 mM at the start of the fermentation runs, dropped upon addition of the feed solution during exponential growth, and remained below 10 mM for the remainder of the run (Figure 15D). (The lowered ammonia levels compared to fermentation run 060327-3 (Figure 14D) are due to the lack of ammonia in the feed solution and reduced ammonia in the base used to maintain the pH. Fermentation runs 060710-4 and 060619-5 showed a spike in ammonia concentration at the end of the runs, but the spikes occurred after the bulk of the production of amorpho-4,11-diene.) The maximum OD₆₀₀ reached was 170 to 220 (Figure 15D), corresponding to 68 g to 88 g DCW/L. The concentration of glucose in the culture medium dropped from 15 g/L to below 1 g/L in less than 20 hours, and remained low (Figure 15E). Acetate levels were low throughout the fermentation runs (Figure 15E).

Example 15

[00299] This example describes the production of amorpho-4,11-diene via the DXP pathway in an *Escherichia coli* host strain.

[00300] Seed cultures of host strains B003, B617, B618, and B619 were established by adding a stock aliquot of each strain to separate 125 mL flasks containing 25 mL M9-MOPS and antibiotics as detailed in Table 1, and by growing the cultures overnight.

[00301] The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05, separate 250 mL flasks containing 40 mL M9-MOPS medium, 45 ug/mL thiamine, micronutrients, 1.00E-5 mol/L FeSO₄, 0.1 M MOPS, 0.5% yeast extract, 20 g/L of D-glucose, and antibiotics. Cultures were incubated at 30°C in a humidified incubating shaker at 250 rpm until they reached an OD₆₀₀ of 0.2 to 0.3, at which point the production of amorpho-4,11-diene in the host cells was induced by adding 40 uL of 1M IPTG to the culture medium.

[00302] At the time of induction, the cultures were overlain with 8 mL of an organic overlay to capture the amorpho-4,11-diene. Samples were taken at various time points, and amorpho-4,11-diene was extracted and analyzed by GC/MS as described in Example 10. Experiments were performed using 2 independent clones of each host strain, and results were averaged. Deviation between samples was found to be less than 10%.

[00303] As shown in Figure 16, *Escherichia coli* host strain B619, which comprises nucleotide sequences encoding enzymes of the full engineered DXP pathway, produced approximately 45 mg/g DCW amorpho-4,11-diene.

Example 16

[00304] This example describes the production of 3-methyl-but-3-en-1-ol and 3-methyl-but-2-en-1-ol in *Escherichia coli* host strains.

- [00305] Seed cultures of host strains B286, B287, B288, and B291 were established by streaking out a stock aliquot of each strain on LB agar containing antibiotics as detailed in Table 1. Three independent colonies were picked for each strain, and each colony was inoculated into 7 mL of LB media containing antibiotics. The cultures were grown overnight at 37°C on a rotary shaker at 250 rpm until late exponential phase. The cultures were then inoculated at an OD₆₀₀ of approximately 0.05, into a 250 mL flask containing 40 mL of M9-MOPS, 2% glucose, 0.5% yeast extract, and antibiotics. The cultures were grown overnight at 37°C on a rotary shaker at 250 rpm until they reached an OD₆₀₀ of approximately 0.2, at which point they were induced by adding 40 µL of 1 M IPTG. The cultures were grown for 72 hours at 30°C on a rotary shaker at 250 rpm. One to two times per day, the OD₆₀₀ of each culture was measured, and a 700 µL sample was removed. To extract the 3-methyl-but-3-en-1-ol and 3-methyl-but-2-en-1-ol from the culture broth, 600 µL of ethyl acetate was added to 300 µL of each removed sample. The sample was then vortexed for 15 minutes, and 400 µL of the upper ethyl acetate phase was transferred to a clean glass vial for analysis.
- [00306] The samples were analyzed on a Hewlett-Packard 6890 gas chromatograph/mass spectrometer (GC/MS). A 1 µL sample was separated on the GC using a DB-5 column (Agilent Technologies, Inc., Palo Alto, CA) and helium carrier gas. The temperature program for the analysis was as follows: 60°C for 3 minutes, increasing temperature at 60°C/minute to a temperature of 300°C, and a hold at 300°C for 2 minutes. The total run time was 9 minutes. The resolved samples were analyzed by a Hewlett-Packard model 5973 mass selective detector. Previous mass spectra demonstrated that 3-methyl-3-buten-1-ol and 3-methyl-2-buten-1-ol have a retention time of 2.067 minutes using this GC protocol. To focus detection on 3-methyl-but-3-en-1-ol and 3-methyl-but-2-en-1-ol, a selective-ion-monitoring method was employed that monitors only ions 56 and 68 in 3-methyl-but-3-en-1-ol and 3-methyl-but-2-en-1-ol.

Example 17

- [00307] This example describes the production of amorpho-4,11-diene by a *Saccharomyces cerevisiae* host strain.
- [00308] The generation of host strain EPY224 is described in Ro *et al.* (*Nature* 440: 940-943; 2006) and in PCT Patent Publication WO2007/005604. Host strain EPY224 was cured of expression plasmid pRS425ADS by growth in YPD medium (Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 ed., ISBN 0-87969-728-8), plating for single colonies on YPD agar, and then patching single colonies onto CSM-Met His agar and CSM-Met Leu agar. Clones that grew on CSM-Met His agar but not on CSM-Met Leu agar were cured (i.e., had lost the plasmid pRS425ADS). One such clone was designated EPY300. EPY300 was transformed with expression plasmid pRS425-ADS-LEU2d, a plasmid identical to pRS425-ADS except that instead of LEU2 it contains a LEU2d selection marker (Erhart and Hollenberg (1983) *J. Bacteriol.* 156: 625-635) yielding host strain Y185.
- [00309] Y185 host cell transformants were selected on synthetic defined media, containing 2% glucose and all amino acids except histidine, leucine, and methionine (CSM-glucose; MP Biomedicals, Solon, OH). The host strain EPY300 is auxotrophic for leucine biosynthesis (*leu2*), but expression plasmid pRS425-ADS-LEU2d in Y185 restores leucine prototrophy (*LEU2*). Single colonies were patched onto selective medium (CSM-glucose-histidine, leucine, methionine), and grown for 2 days. The cells were scraped from the plate and transferred to 1 mL of 25% (v/v) glycerol in a cryotube. The suspension was mixed, and then stored at -80°C.
- [00310] Seed flasks of host strain Y185 were established by adding a stock aliquot of the strain to a 125 mL flask containing 25 mL of CSM-glucose lacking leucine and methionine, and by growing the cultures overnight. The cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05 a 250 mL baffled flask

containing 40 mL of synthetic defined media lacking leucine, and containing 0.2% glucose, 1.8% galactose, and 1 mM methionine. The culture was incubated at 30°C on a rotary shaker at 200 rpm. Because the presence of glucose in the media prevents induction of the *GALI* promoter by galactose, amorpho-4,11-diene production was not induced until the cells had used up the glucose in the media and had switched to using galactose as their main carbon source. At the time of inoculation, the cultures were overlain with 8 mL of an organic overlay to capture the amorpho-4,11-diene. Samples were taken at 72 hours by transferring 5 uL of the organic solvent layer to a clean glass vial containing 500 uL ethyl acetate containing a known concentration of beta- or trans-caryophyllene as an internal standard.

[00311] The organic overlay/ethyl acetate samples were analyzed on a Hewlett-Packard 6890 gas chromatograph/mass spectrometer (GC/MS) as described in Example 10.

[00312] After 72 hours of growth, 3 yeast cultures were found to produce 60.68, 54.48, and 59.25 mg/L amorpho-4,11-diene.

Example 18

[00313] This example describes the production of amorpho-4,11-diene in an *Saccharomyces cerevisiae* host strain where the host strain includes a native mevalonate pathway as well as a heterologous mevalonate pathway that is under control of a heterologous regulatory control.

[00314] Yeast strains CEN.PK2-1C (Y002) (MATA; ura3-52; trp1-289; leu2-3,112; his3Δ1; MAL2-8C; SUC2) and CEN.PK2-1D (Y003) (MATα; ura3-52; trp1-289; leu2-3,112; his3Δ1; MAL2-8C; SUC2) (J. P. van Dijken *et al.*, *Enzyme Microb Technol* 26, 706 (Jun 1, 2000) were cultivated in either standard rich medium (YPD) or in defined synthetic medium (J. D. Rose, F. Winston, P. Heiter, *Methods in yeast genetics: a laboratory course manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1990) lacking appropriate nutrients allowing for selection of integrative transformants, plasmid retention, and meiotic progeny.

[00315] DNA-mediated transformations into *S. cerevisiae* were conducted using the lithium acetate procedure as described by R. H. Schiestl, R. D. Gietz, *Curr Genet* 16, 339 (Dec, 1989). All gene disruptions and replacements were confirmed by phenotypic analysis, colony polymerase chain reaction ("PCR") and sequencing of amplified genomic DNA. Plasmids pAM489-pAM498 were constructed using the pCR 2.1 (Invitrogen, Carlsbad CA) and are schematically described by Figure 7A-C and Table 6. The HISMX marker sequences are described in M. S. Longtine *et al.*, *Yeast* 14, 953 (Jul, 1998). Propagation of plasmid DNA was performed in *Escherichia coli* strain DH5α.

Table 6

Strain	5'HR	Gene #1	Crick Promoter	Watson Promoter	Gene #2	Genetic Marker	3'HR
pAM489	TRP1	tHMGR	GAL1	GAL10	ERG20	TRP1	TRP1
pAM490	TRP1	tHMGR	CUP1	CUP1	ERG20	TRP1	TRP1
pAN491	URA3	tHMGR	GAL1	GAL10	ERG13	URA3	URA3
pAM492	URA3	IDH1	CUP1	CUP1	tHMGR	URA3	URA3
pAM493	ADE1	tHMGR	GAL1	GAL10	IDH1	ADE1	URA3
pAM494	ADE1	tHMGR	CUP1	CUP1	IDH1	ADE1	ADE1
pAM495	HIS3	ERG12	GAL1	GAL10	ERG10	HISMX	HIS3
pAM496	HIS3	ERG12	CUP1	CUP1	ERG10	HISMX	HIS3
pAM497	LEU2	ERG19	GAL1	GAL1	ERG8	HISMX	LEU2
pAM498	LEU2	ERG19	CUP1	CUP1	ERG8	HISMX	LEU2

[00316] *S. cerevisiae* strains Y002 and Y003 were prepared for introduction of inducible mevalonate pathway genes by the following. The *ERG9* promoter was replaced with the *S. cerevisiae* *MET3* promoter by PCR

amplification of the KanMX-PMET3 region from pAM328 (SEQ ID NO: 43) using primers 50-56-pw100-G (SEQ ID NO: 44) and 50-56-pw101-G (SEQ ID NO: 45) containing 45 basepairs of homology to the native *ERG9* promoter. 10µg of the resulting PCR product was transformed into exponentially growing Y002 and Y003 strains using 40% w/w polyethylene glycol 3350 (Sigma-Aldrich St Louis, MO), 100 mM lithium acetate (Sigma), 10µg Salmon Sperm DNA (Invitrogen) and incubation at 30°C for 30 minutes followed by a 42°C heat shock for 30 minutes (as described by Schiestl & Gietz, *Curr. Genet.* 16: 339 (1989)). Positive recombinants were identified by their ability to grow on rich medium containing 0.5µg/ml Geneticin (Invitrogen Co, Carlsbad, CA) and confirmed by diagnostic PCR. The resultant clones were given the designation Y93 (MAT A) and Y94 (MAT alpha). Next, the *ADE1* open reading frame was replaced with the *Candida glabrata LEU2* gene (*CgLEU2*). The 3.5KB *CgLEU2* genomic locus was amplified from *C. glabrata* genomic DNA (ATCC, Manassas, VA) using primers 61-67-CPK066-G (SEQ ID NO: 46) and 61-67-CPK067-G (SEQ ID NO: 47) containing 50 basepairs of flanking homology to the *ADE1* open reading frame (ORF). 10µg of the resulting PCR product was transformed into exponentially growing Y93 and Y94 as described above. *ade1*- strains were selected for growth in the absence of leucine supplementation and confirmed by diagnostic PCR. The resultant clones were given the designation Y176 (MAT A) and Y177 (MAT alpha).

- [00317] To generate *S. cerevisiae* strain Y188, 2µg's of plasmid DNA from pAM491 (SEQ ID NO: 48) and pAM495 (SEQ ID NO: 49), respectively, were digested overnight with PmeI (New England Biolabs, Beverly, MA) and introduced into exponentially growing Y176 as described above. Positive recombinants were selected for by growth on medium lacking uracil and histidine. Integration into the correct genomic locus was confirmed by diagnostic PCR.
- [00318] To generate *S. cerevisiae* strain Y189, 2µg's of plasmid DNA from pAM489 (SEQ ID NO: 50) and pAM497 (SEQ ID NO: 51), respectively, were digested overnight with PmeI and introduced into exponentially growing Y177 as described above. Positive recombinants were selected for by growth on medium lacking tryptophan and histidine. Integration into the correct genomic locus was confirmed by diagnostic PCR.
- [00319] Approximately 1×10^7 cells from Y188 and Y189 were mixed on a YPD medium plate for 6 hours at room temperature to allow for mating. The mixed cell culture was then plated to medium lacking histidine, uracil and tryptophan to select for growth of diploid cells. 2µg of plasmid DNA from pAM493 (SEQ ID NO: 52) was digested overnight with PmeI and introduced into exponentially growing diploid cells as described above. Positive recombinants were selected for by growth on medium lacking adenine. Integration into the correct genomic locus was confirmed by diagnostic PCR. The resultant strain was given the designation Y238.
- [00320] To generate haploid strains containing the full complement of introduced genes, Y238 was sporulated in 2% potassium acetate and 0.02% raffinose liquid medium. Approximately 200 genetic tetrads (tetrads are four-spored meiotic products) were isolated using a Singer Instruments MSM300 series micromanipulator (Singer Instrument Co, LTD. Somerset, UK). Independent genetic isolates containing the appropriate complement of introduced genetic material were identified by their ability to grow in the absence of adenine, histidine, uracil, and tryptophan. Integration of all introduced DNA was confirmed by diagnostic PCR. The resultant strains were given the designation Y210 (MAT A) and Y211 (MAT alpha).
- [00321] 2µg of plasmid DNA from pAM426 (SEQ ID NO: 53), containing *S. cerevisiae* condon optimized Amorphadine Synthase (ADS) expressed from the *S. cerevisiae GAL1* promoter, was introduced into

exponentially growing Y210 and Y211 as described above. *S. cerevisiae* strains that contained the pAM426 plasmid were selected for by their ability to grow in the absence of leucine supplementation. The resultant strains were given the designation Y225 (MAT A) and Y227 (MAT alpha).

[00322] 2 μ g of plasmid DNA from pAM322 (SEQ ID NO: 54), containing *S. cerevisiae* codon optimized Amorphadine Synthase (ADS) and cytochrome P450 monooxygenase (AMO) expressed from the *S. cerevisiae* *GALI* and the cytochrome P450 oxidoreductase (CPR) expressed from the *S. cerevisiae* *GAL10* promoter, was introduced into exponentially growing Y210 and Y211 as described above. *S. cerevisiae* strains that contained the pAM322 plasmid were selected for by their ability to grow in the absence of leucine supplementation. The resultant strains were given the designation Y222 (MAT A) and Y224 (MAT alpha).

Example 19

[00323] This example describes the production of α -farnesene or β -farnesene in *Escherichia coli* host strains.

[00324] Seed cultures of host strains B552 and B592 were established by adding a stock aliquot of each strain to a 125 mL flask containing 25 mL M9-MOPS, 0.8% glucose, 0.5% yeast extract, and antibiotics as detailed in Table 1, and by growing the cultures overnight.

[00325] The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05, 250 mL flasks containing 40 mL M9-MOPS, 2% glucose, 0.5% yeast extract, and antibiotics. Cultures were incubated at 30°C on a rotary shaker at 250 rpm until they reached an OD₆₀₀ of approximately 0.2, at which point the production of α -farnesene or β -farnesene in the host cells was induced by adding 40 μ L of 1 M IPTG. At the time of induction, the cultures were overlain with 8 mL of an organic overlay to capture the α -farnesene. Samples were taken every 24 hours up to 120 hours (total of 5 time points) by transferring 2 μ L to 10 μ L of the organic overlay layer to a clean glass vial containing 1 mL ethyl acetate spiked with trans-caryophyllene as an internal standard. In addition, 1 mL aliquots of the cultures were spun down, cell pellets were resuspended in 250 μ L sterile water, and the cell suspensions were transferred to a glass vial containing 1 mL ethyl acetate spiked with trans-caryophyllene as an internal standard. In addition, 0.5 mL aliquots of the whole culture broth were added to a glass vials containing 1 mL ethyl acetate spiked with trans-caryophyllene as an internal standard. The whole culture broth samples were extracted in the ethyl acetate by vortexing the glass vials for 10 minutes, after which 600 μ L of the ethyl acetate extraction was transferred to a clean glass vial.

[00326] The organic overlay/ethyl acetate samples and the ethyl acetate-extracted whole culture broth samples were analyzed on an Agilent 6890N gas chromatograph equipped with an Agilent 5975 mass spectrometer (GC/MS) in full scan mode (50-500 m/z). To expedite run times, the temperature program and column matrix was modified to achieve optimal peak resolution and the shortest overall runtime. A 1 μ L sample was separated using a HP-5MS column (Agilent Technologies, Inc., Palo Alto, CA) and helium carrier gas. The temperature program for the analysis was as follows: 150°C hold for 3 minutes, increasing temperature at 25°C/minute to a temperature of 200°C, increasing temperature at 60°C/minute to a temperature of 300°C, and a hold at 300°C for 1 minute. Previous mass spectra demonstrated that the β -farnesene synthase product was β -farnesene, and that β -farnesene had a retention time of 4.33 minutes using this GC protocol. Farnesene titers were calculated by comparing generated peak areas against a quantitative calibration curve of purified β -farnesene (Sigma-Aldrich Chemical Company, St. Louis, MO) in trans-caryophyllene-spiked ethyl acetate.

[00327] Host strain B592 produced approximately 400 mg/L of α -farnesene at 120 hours (averaged over 3 independent clones), and had a maximal specific productivity of approximately 46 mg/L/OD₆₀₀. Host strain B552 produced approximately 1.1 g/L of β -farnesene at 120 hours (averaged over 3 independent clones), and had a maximal specific productivity of approximately 96 mg/L/OD₆₀₀ (1 representative clone).

Example 20

[00328] This example describes the production of β -farnesene via the DXP pathway in an *Escherichia coli* host strain.

[00329] Seed cultures of host strains B650, B651, B652, and B653 were established by adding a stock aliquot of each strain to separate 125 mL flasks containing 25 mL M9-MOPS and antibiotics as detailed in Table 1, and by growing the cultures overnight.

[00330] The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05 separate 250 mL flasks containing 40 mL M9-MOPS minimal medium, 45 ug/mL thiamine, micronutrients, 1.00E-5 mol/L FeSO₄, 0.1 M MOPS, 0.5% yeast extract, 20 g/L of D-glucose, and antibiotics. The cultures were incubated at 30°C in a humidified incubating shaker at 250 rpm until they reached an OD₆₀₀ of 0.2 to 0.3, at which point the production of β -farnesene in the host cells was induced by adding 40 uL of 1 M IPTG to the culture medium. At the time of induction, the cultures were overlain with 8 mL of an organic overlay to capture the β -farnesene. Samples were taken at various time points by transferring 100 uL samples of the upper organic overlay layer to a clean tube. The tube was centrifuged to separate out any remaining cells or media, and 10 uL of the organic overlay samples were transferred into 500 uL ethyl acetate spiked with beta- or trans-caryophyllene as an internal standard in clean glass GC vials. The mixtures were vortexed for 30 seconds, and then analyzed as described in Example 18. *Escherichia coli* host strain B653 produced approximately 7 mg/g DCW β -farnesene.

Example 21

[00331] This example describes the production of α -farnesene or β -farnesene in a *Saccharomyces cerevisiae* host strain.

[00332] Strain EPY300 was generated by removing the expression plasmid from *Saccharomyces cerevisiae* strain EPY224 (Ro *et al.* (2006) *Nature* 440: 940-943; PCT Patent Publication WO2007/005604) by culturing in rich medium. Strain EPY300 was then transformed with expression plasmids pRS425-FSA or pR425-FSB, yielding host strains Y166 and Y164, respectively.

[00333] Host cell transformants were selected on synthetic defined media, containing 2% glucose and all amino acids except leucine (SM-glu). The host strain EPY300 was auxotrophic for leucine biosynthesis (*leu2*), but expression plasmid pRS425-FSA or pRS425-FSB restores leucine prototrophy (*LEU2*). Single colonies were transferred to culture vials containing 5 mL of liquid SM-glu lacking leucine. The cultures were incubated by shaking at 30°C until growth reaches stationary phase. The cells were stored at -80°C in cryovials in 1 mL frozen aliquots made up of 400 μ L 50% glycerol and 600 μ L liquid culture.

[00334] Seed cultures were established by adding a stock aliquot to a 125 mL flask containing 25 mL SM-glu lacking leucine, and growing the cultures overnight. The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05 250 mL baffled flasks containing 40 mL of synthetic defined media lacking leucine, 0.2% glucose, and 1.8% galactose. Cultures were incubated at 30°C on a rotary shaker at 200 rpm. Because the presence of glucose in the media prevents induction of the *Gall* promoter by galactose, farnesene production was not induced until the cells use up the glucose in the media and switch to using galactose as their main carbon source. The cultures are overlain with 8 mL methyl oleate or isopropyl

myristate. Samples were taken once every 24 hours by transferring 2-10 uL of the organic solvent layer to a clean glass vial containing 500 uL ethyl acetate containing a known concentration of beta- or trans-caryophyllene as an internal standard. In addition, 0.5 mL aliquots of the whole culture broth were added to a glass vials containing 1 mL ethyl acetate spiked with trans-caryophyllene as an internal standard. The whole culture broth samples were extracted in the ethyl acetate by vortexing the glass vials for 10 minutes, after which 600 uL of the ethyl acetate extraction was transferred to a clean glass vial.

[00335] Host strain Y166 produced approximately 9.8 mg/L of α -farnesene at 120 hours (averaged over 3 independent clones), and had a maximal specific productivity of approximately 3 mg/L/OD₆₀₀ (1 representative clone). Host strain Y164 produced approximately 56 mg/L of β -farnesene at 120 hours (averaged over 3 independent clones), and had a maximal specific productivity of approximately 20 mg/L/OD₆₀₀ (1 representative clone).

Example 22

[00336] This example describes the production of γ -terpinene, α -pinene, and terpinolene in *Escherichia coli* host strains.

[00337] Seed cultures of host strains for production of γ -terpinene (*E. coli* DH1-T1r [pMevT, pMevB-Gpps, pAM445]), α -pinene (*E. coli* DH1-T1r [pMevT, pMevB-Gpps, pAM443 or pAM442]) or terpinolene (*E. coli* DH1-T1r [pMevT, pMevB-Gpps, pAM444]) were established by adding a stock aliquot of each strain to separate 125 mL flasks containing 25 mL M9-MOPS, 2% glucose, 0.5% yeast extract, and antibiotics as detailed in Table 1, and by growing the cultures overnight to late exponential phase.

[00338] The seed cultures were used to inoculate at an initial OD₆₀₀ of approximately 0.05, 250 mL flasks containing 40 mL M9-MOPS, 2% glucose, 0.5% yeast extract, and antibiotics. At time of inoculation, the cultures were also overlain with 4 mL hexadecane. Cultures were incubated at 30°C on a rotary shaker at 200 - 250 rpm until they reached an OD₆₀₀ of approximately 0.2, at which point the production of the compound of interest in the host cells in the host cells was induced by adding 40 uL of 1 M IPTG. Samples were taken once per day for 96 hours by transferring 200 uL of the hexadecane layer to a 0.6 mL microfuge tube. For analysis, the hexadecane overlay was diluted 1:1 or 1:10 with ethyl acetate spiked with trans-caryophyllene as an internal standard in a 1.8 mL GC vial. In addition, 1 mL aliquots of the cultures were spun down, cell pellets were resuspended in 250 uL sterile water, and the cell suspensions were transferred to a glass vial containing 1 mL ethyl acetate spiked with trans-caryophyllene as an internal standard. The cell pellets were extracted in the ethyl acetate by vortexing the glass vials for 15 minutes, after which 500 uL of the ethyl acetate extraction was transferred to a clean glass vial.

[00339] The hexadecane/ethyl acetate samples and the ethyl acetate-extracted cell pellet samples were analyzed on an Agilent 6890N gas chromatograph equipped with an Agilent 5975 mass spectrometer (GC/MS) in full scan mode (50-500 m/z). To expedite run times, the temperature program and column matrix was modified to achieve optimal peak resolution and the shortest overall runtime. A 1 μ L sample was split (a split ratio between 1:2 and 1:50 was selected based on sample concentration) and then separated using a HP-5MS column (Agilent Technologies, Inc., Palo Alto, CA) and helium carrier gas. The temperature program for the analysis was as follows: 75°C hold for 3 minutes, increasing temperature at 20°C/minute to a temperature of 115°C, increasing temperature at 60°C/minute to a temperature of 300°C, and a hold at 300°C for 0.5 minute. The various products, γ -terpinene, α -pinene, and terpinolene were observed at 5.4, 4.1, 5.4, and 5.9 minutes, respectively. Titters were calculated by comparing generated peak areas against a quantitative calibration curve of purified standards in trans-caryophyllene-spiked ethyl acetate.

Example 23

[00340] This example describes the production of linalool, limonene, β -pinene, β -phellandrene, carene, or sabinine in *Escherichia coli* host strains.

[00341] Seed cultures are established by adding a stock aliquot of each strain to separate 125 mL flasks containing 25 mL M9-MOPS, 0.5% yeast extract, 2% glucose, and antibiotics as detailed in Table 1, and by growing the cultures overnight.

[00342] The seed cultures are used to inoculate at an initial OD₆₀₀ of approximately 0.05, 250 mL baffled flasks containing 40 mL M9-MOPS, 0.5% yeast extract, 2% glucose, and antibiotics. Cultures are incubated at 30°C on a rotary shaker at 250 rpm until they reach an OD₆₀₀ of approximately 0.2, at which point the production of the compound of interest in the host cells is induced by adding 40 μ L of 1 M IPTG to the culture medium. The compound of interest is separated from the culture medium through solvent-solvent extraction, or by settling and decantation if the titer of the compound of interest is large enough to saturate the media and to form a second phase.

Sequence Listing

SEQ ID NO: 1

MevT66 operon

GAATTCAAAGGAGGAAAATAAAATGAAGAACTGTGTGATTGTTTCTGCGGTCCGCACGGCGATCGGC
AGCTTTAACGGCTCTTTAGCGAGCACCTCTGCAATCGATCTGGGTGCGACGGTCATTAAGGCCGCCATT
GAACGCGCCAAAATCGACAGCCAGCACGTTGATGAGGTGATCATGGGCAATGTGTTACAAGCCGGCC
TGGGTCAAAACCCAGCGCGTCAAGCACTGTATAAAATCTGGTCTGGCCGAGACCGTGTGTGGCTTCACC
GTCAATAAGGTTTGCGGCTCTGGCCTGAAGAGCGTGGCCCTGGCAGCACAAAGCGATTCAAGCCGGTCA
GGCACAAAGCATCGTTGCGGGTGGCATGGAGAACATGTCTCTGGCGCCGTACTTATTAGATGCCAAAG
CCCGCAGCGGTTATCGCCTGGGCGATGGTCAGGTGTACGACGTCATCTTACGCGATGGCTTAATGTGC
GCGACCCACGGTTACCACATGGGTATTACGGCCGAAAACGTGGCGAAAGAATACGGCATTACGCGCG
AGATGCAGGATGAATTAGCACTGCACTCTCAGCGCAAAGCAGCAGCCGCGATCGAGTCTGGTGCGTTT
ACGGCGGAAATCGTGCCAGTTAACGTGGTCACGCGCAAGAAGACGTTTCGTTTTACGCCAGGACGAGTT
CCCGAAGGCAAACAGCACCGCGGAGGCCTTAGGTGCCTTACGCCAGCCTTTGACAAAGCGGGGCACG
GTCACCGCCGGTAATGCGAGCGGCATCAATGATGGTGCAGCGGCACTGGTCATCATGGAAGAGAGCG
CCGCATTAGCAGCGGGTCTGACCCCATTAGCGCGCATTAAATCTTATGCCAGCGGCGGGCGTCCCACCA
GCCCTGATGGGCATGGGTCCGGTCCCAGCCACGCAAAAAGCCCTGCAATTAGCGGGCCTGCAACTGGC
CGACATTGATCTGATCGAGGCGAACGAGGCGTTTGACGCGCAGTTTCCTGGCGGTGGGTAAGAATCTGG
GCTTCGACAGCGAGAAAAGTCAATGTGAACGGTGGCGCGATTGCGTTAGGCCATCCGATTGGTGCAAGC
GGCGCACGCATCTTAGTGACGTTACTGCACGCCATGCAGGCACGCGACAAGACCTTAGGCCTGGCGAC
CTTATGTATTGGTGGCGGTCAAGGTATCGCCATGGTGATCGAACGCCCTGAACTGAAGATCTAGGAGGA
AAGCAAAATGAACTGAGCACCAAGCTGTGCTGGTGTGGCATCAAGGGTCGCCTGCGCCACAAAAG
CAGCAACAGCTGCACAACACGAACCTGCAAATGACCGAGCTGAAAAAGCAGAAGACGGCCGAGCAA
AAGACCCGCCCCGAGAACGTTGGCATCAAGGGCATCCAGATTTATATCCCGACGCAGTGTGTCAACCA
ATCTGAGCTGGAGAAATTCGATGGCGTCAGCCAGGGTAAGTACACCATCGGCCTGGGCCAGACCAAC
ATGAGCTTCGTGAACGACCGTGAGGACATCTATTCTATGAGCCTGACGGTGCTGTCTAAGCTGATCAA
GAGCTACAACATCGACACGAATAAGATCGGTCGTCTGGAGGTGGGTACGGAGACGCTGATTGACAAG
AGCAAAAGCGTGAAGTCTGTCTTAATGCAGCTGTTTCGGCGAGAACACGGATGTGAGGGTATCGACAC
CCTGAACGCGTGTTACGGCGGCACCAACGCACTGTTCAATAGCCTGAACTGGATTGAGAGCAACGCCT
GGGATGGCOGCGATGCGATCGTGTGCGGCGATATCGCCATCTATGACAAGGGTGCGGCACGTCGG
ACCGGCGGTGCAGGCACCGTTGCGATGTGGATTGGCCCGGACGCACCAATTGTCTTCGATTCTGTCCG
CGCGTCTTACATGGAGCACGCCTACGACTTTTACAAGCCGGACTTCACGAGCGAATACCCGTACGTGG
ACGGCCACTTCTCTCTGACCTGCTATGTGAAGGCGCTGGACCAGGTTTATAAGTCTTATAGCAAAAAG
GCGATTTCTAAGGGCCTGGTCAGCGACCCGGCAGGCAGCGACGCCCTGAACGTGCTGAAGTATTTCTGA
CTACAACGTGTTCCATGTCCCGACCTGCAAATTAGTGACCAAATCTTATGGCCGCCTGTTATATAATGA
TTTCCGTGCCAACCCGCGAGCTGTTCCCGGAGGTTGACGCCGAGCTGGCGACGCGTGATTACGACGAGA
GCCTGACCGACAAGAACATCGAGAAGACCTTCGTCAACGTGCGGAAGCCGTTCCACAAAGAGCGTGT
GGCCCAAAGCCTGATCGTCCCGACCAACACGGGCAACATGTATACCGCGTCTGTCTACGCGGCATTTCG
CGAGCCTGCTGAATTACGTGCGTTCTGACGACCTGCAGGGCAAGCGCGTTGGCCTGTTACGCTACGGT
AGCGGCTTAGCGGCCAGCCTGTATAGCTGCAAAATTGTGCGCGACGTCCAGCACATCATCAAGGAGCT

GGACATCACCACAAAGCTGGCGAAGCGCATCACCGAGACGCCGAAAGATTACGAGGCAGCGATCGAG
TTACGCGAGAATGCGCATCTGAAGAAGAACTTCAAGCCGCAAGGTAGCATCGAGCACCTGCAGAGCG
GCGTCTACTACCTGACGAACATTGACGACAAGTTCCGCCGTTTCTTATGACGTCAAAAAGTAACTAGTA
GGAGGAAAACATCATGGTGCTGACGAACAAAACCGTCATTAGCGGCAGCAAGGTGAAGTCTCTGAGC
AGCGCCCCAAAGCTCTAGCAGCGGCCCGTCTAGCAGCAGCGAGGAGGACGACAGCCGTGACATTGAGT
CTCTGGACAAGAAGATCCGCCCGCTGGAGGAGTTAGAGGCCCTGCTGAGCAGCGGCAACACCAAGCA
GCTGAAGAACAAGGAAGTTGCAGCGCTGGTGATCCACGGTAAGCTGCCACTGTATGCGCTGGAAAAG
AACTGGGCGATACGACGCGTGCGGTCGCGGTGCGTCGCAAAGCCTTAAGCATCTTAGCGGAGGCCCC
GGTGTTAGCCAGCGACCGCCTGCCGTACAAGAACTACGACTACGACCGCGTGTTTGGCGCGTGCTGCG
AGAATGTCATTGGCTACATGCCGTTACCGGTTGGTGATCGGCCCGCTGGTCATTGATGGCACGAGC
TATCACATTCCAATGGCGACCACGGAAGGTTGCTTAGTCGCCAGCGCCATGCGTGGCTGTAAGGCGAT
TAACGCCGCGCGTGGCGCGACGACCGTGTTAACCAAGGATGGTATGACGCGCGGTCCGGTCGTCCGCT
TCCCAACGCTGAAGCGCAGCGGCGCGTGTAAGATTTGGCTGGATTCTGAGGAGGGCCAAAACGCGAT
CAAGAAAGCCTTCAACTCTACGAGCCGTTTCGCGCGTTTACAGCATATCCAGACCTGCCTGGCCGGCG
ACCTGCTGTTTCATGCGCTTCCGCACCACCACGGGCGATGCGATGGGCATGAACATGATCAGCAAGGGC
GTCGAATATAGCCTGAAACAAATGGTGGAAGAATATGGCTGGGAGGACATGGAGGTTGTCTCTGTGA
GCGGCAACTATTGCACCGACAAGAAGCCGGCAGCCATTAAGTGGATTGAGGGTCGCGGCAAAAGCGT
CGTGGCAGAAGCGACCATCCCAGGCGACGTGGTCCGTAAGGTTCTGAAGAGCGACGTCAGCGCCCTG
GTTGAGTTAAATATCGCGAAAAACCTGGTCGGCAGCGCGATGGCGGGCAGCGTGGGTGGCTTTAACGC
ACATGCAGCGAATCTGGTTACGGCGGTTTTCTTAGCCTTAGGTCAGGACCCAGCCCCAAAATGTCGAGA
GCAGCAACTGCATTACCTTAATGAAAGAGGTTGACGGTGACCTGCGCATCAGCGTTTCTATGCCGTCT
ATCGAGGTCGGCACGATCGGCGGGCGGCACCGTTTTAGAACCGCAAGGTGCGATGCTGGATCTGCTGGG
CGTGCGCGGCCACATGCAACGGCCCCAGGCACCAATGCCCGCCAACTGGCCCGTATCGTGGCCTGCG
CGGTTCTGGCGGGTGAGCTGAGCCTGTGCGCCGCATTAGCCGCGGGCCATTTAGTTCAATCTCACATG
ACCCACAACCGCAAGCCGGCAGAACCAACCAAGCCAAATAACCTGGACGCAACCGACATTAACCGTC
TGAAGGATGGCAGCGTCACGTGCATTAAAAGCTGAGCATGCTACTAAGCTT

SEQ ID NO: 2

Primer 4-49 mvaA SpeI

5'-GCTACTAGTAGGAGGAAAACATCATGCAAAGTTTAGATAAGAATTTCCG-3'

SEQ ID NO: 3

Primer 4-49 mvaAR XbaI

5'-GCTTCTAGACTATTGTTGTCTAATTTCTTGTAATAATGCG-3'

SEQ ID NO: 4

Primer HMGS 5' Sa mvaS-S

5'-GAACTGAAGATCTAGGAGGAAAAGCAAAATGACAATAGGTATCGACAAAATAAACT-3'

SEQ ID NO: 5

Primer HMGS 3' Sa mvaS-AS

5'-TTGCATGATGTTTTCTCTACTAGTTACTCTGGTCTGTGATATTCGCGAAC-3'

SEQ ID NO: 6

Primer 19-25 atoB SfiI-S

5'-GCTAGGCCATCCTGGCCATGAAGAACTGTGTGATTGTTTCTG-3'

SEQ ID NO: 7

Primer 19-25 mvaA-AsiSI-AS

5'-GCTTGCGATCGCCGGCGGATTTGTCCTACTCAG-3'

SEQ ID NO: 8

Primer 9-70C

5'-CCACCTCGAGATGTCATTACCGTTCCTAACCTCTG-3'

SEQ ID NO: 9

Primer 26-39B

5'-TGGTGGAGCTCTTATTTAAGCTGGGTAAATGCAGATAATCG-3'

SEQ ID NO: 10

Primer 26-39A

5'-TTCTTGAGCTCTTATTCCTTTGGTAGACCAGTCTTTGCG-3'

SEQ ID NO: 11

Primers 4-40 mvaEF BamHI

5' – TATGGATCCTAAGGAGGATATTTAGATGAAAACAGTAGTTATTATTGATGC – 3'

SEQ ID NO: 12

Primer 4-40 mvaER HindIII

5'- AGCTAAGCTTTTATTGTTTTCTTAAATCATTTAAAATAGC – 3'

SEQ ID NO: 13

Primer 4-40 mvaSF BglII

5' – TATAGATCTTAAGGAGGATATTTAGATGACAATTGGGATTGATAAAAATTAG – 3'

SEQ ID NO: 14

Primer 4-39 mvaSR BamHI

5' – TTTGGATCCTTAGTTTTCGATAAGAGCGAACGG – 3'

SEQ ID NO: 15

Primer 67-1A-C for PCR amplification of the coding sequence of the dxs gene

5'- ACA CTC GAG GAG GAA TAA ATG AGT TTT GAT ATT GCC AAA TAC CCG -3'

SEQ ID NO: 16

Primer 67-1B-C for PCR amplification of the coding sequence of the dxs gene

5'- TGA TGG TAC CTT ATG CCA GCC AGG CCT TGA TTT TGG C -3'

SEQ ID NO: 17

Primer 67-1C-C for PCR amplification of the coding sequence of the dxr gene

5'- ACT AGG TAC CAG GAG GAA TAA ATG AAG CAA CTC ACC ATT CTG GGC -3'

SEQ ID NO: 18

Primer 67-1D-C for PCR amplification of the coding sequence of the dxr gene

5'- AAT TGA TGG GCC CTC AGC TTG CGA GAC GCA TCA CCT C -3'

SEQ ID NO: 19

Primer 67-1E-C for PCR amplification of the coding sequence of the ispD gene

5'- CAT AAA GGG CCC AGG AGG AAT AAA TGG CAA CCA CTC ATT TGG ATG -3'

SEQ ID NO: 20

Primer 67-1F-C for PCR amplification of the coding sequence of the ispD gene

5'- TAT TGT TCA TAT GTT ATG TAT TCT CCT GAT GGA TGG TTC G -3'

SEQ ID NO: 21

Primer 67-1G-C for PCR amplification of the coding sequence of the ispE gene

5'- AAC TAA CAC ATA TGA GGA GGA ATA AAT GCG GAC ACA GTG GCC CTC -3'

SEQ ID NO: 22

Primer 67-1H-C for PCR amplification of the coding sequence of the ispE gene

5'- TGT TAG TTA CGC GTT TAA AGC ATG GCT CTG TGC AAT GG -3'

SEQ ID NO: 23

Primer 67-2A-C for PCR amplification of the coding sequence of the ispF gene

5'- ACG GGA TCC AGG AGG AAT AAA TGC GAA TTG GAC ACG GTT TTG ACG -3'

SEQ ID NO: 24

Primer 67-2B-C for PCR amplification of the coding sequence of the ispF gene

5'- TTT AGT TGG GCC CTC ATT TTG TTG CCT TAA TGA GTA GCG CC -3'

SEQ ID NO: 25

Primer 67-2C-C for PCR amplification of the coding sequence of the ispG gene

5'- TAC TAA GGG CCC AGG AGG AAA TAA TGC ATA ACC AGG CTC CAA TTC AAC G -3'

SEQ ID NO: 26

Primer 67-2D-C for PCR amplification of the coding sequence of the ispG gene

5'- TCC GGG TAC CTT ATT TTT CAA CCT GCT GAA CGT CAA TTC G -3'

SEQ ID NO: 27

Primer 67-2E-C for PCR amplification of the coding sequence of the *ispH* gene

5'- AAC AGG TAC CAG GAG GAA ATA ATG CAG ATC CTG TTG GCC AAC C -3'

SEQ ID NO: 28

Primer 67-2F-C for PCR amplification of the coding sequence of the *ispH* gene

5'- TGG ATG AAG TCG ACT TAA TCG ACT TCA CGA ATA TCG ACA CGC AGC -3'

SEQ ID NO: 29

Primer 67-2G-C for PCR amplification of the coding sequence of the *idi* gene

5'- CAT CAA GTC GAC AGG AGG AAA TAA TGC AAA CGG AAC ACG TCA TTT TAT TG -3'

SEQ ID NO: 30

Primer 67-2H-C for PCR amplification of the coding sequence of the *idi* gene

5'- TAA TGC AAG CTT ATT TAA GCT GGG TAA ATG CAG ATA ATC G -3'

SEQ ID NO: 31

Primer 67-2I-C for PCR amplification of the coding sequence of the *ispA* gene

5'- CAG TAA AGC TTA GGA GGA AAT AAT GGA CTT TCC GCA GCA ACT CG -3'

SEQ ID NO: 32

Primer 67-2J-C for PCR amplification of the coding sequence of the *ispA* gene

5'- TAG TTC CAT GGT TAT TTA TTA CGC TGG ATG ATG TAG TCC GC -3'

SEQ ID NO: 33

Primer 9-156A for PCR amplification of the *RK2 par* locus

5'- ACATAGACGTCGGGAAAGCGAGGATCTAGGTAGGG -3'

SEQ ID NO: 34

Primer 9-156B for PCR amplification of the *RK2 par* locus

5'- TTCCCGCTCGAGGTGGCGGACCATATAGGCAGATCAG -3'

SEQ ID NO: 35

Primer 19-137 *cml*-pAM37-AS

5' – GACGTCGATATCTGGCGAAAATG – 3'

SEQ ID NO: 36

Primer 19-137 *cml*-pAM37-S

5' – TACTAGTGCTTGGATTCTCACC – 3'

SEQ ID NO: 37

Primer for PCR amplification of a nucleotide sequence encoding a β -farnesene synthase

5'-CCATGGACACTCTGCCGATCTCTTCCGTAAGC-3'

SEQ ID NO: 38

Primer for PCR amplification of a nucleotide sequence encoding a β -farnesene synthase

5'-GAGCTCTCATACGACCATAGGGTGTACG-3'

SEQ ID NO: 39

Primer for PCR amplification of a nucleotide sequence encoding an α -farnesene synthase

5'-CCATGGACCTGGCAGTAGAAATTGC-3'

SEQ ID NO: 40

Primer for PCR amplification of a nucleotide sequence encoding an α -farnesene synthase

5'-GAGCTCTTACATCGGTACCGGCTCCAG-3'

SEQ ID NO: 41

atoB(opt):*HMGS*(opt):*mvaA* operon

ATGAAGAACTGTGTGATTGTTTCTGCGGTCCGCACGGCGATCGGCAGCTTTAACGGCTCTTTAGCGAG
CACCTCTGCAATCGATCTGGGTGCGACGGTCATTAAGGCCGCCATTGAACGCGCCAAAATCGACAGCC
AGCACGTTGATGAGGTGATCATGGGCAATGTGTTACAAGCCGGCCTGGGTCAAACCCAGCGCGTCAA
GCACTGTTAAAATCTGGTCTGGCCGAGACCGTGTGTGGCTTCACCGTCAATAAGGTTTTCGGCTCTGGC
CTGAAGAGCGTGGCCCTGGCAGCACAAGCGATTCAAGCCGGTCAGGCACAAAGCATCGTTGCGGGTG
GCATGGAGAACATGTCTCTGGCGCCGTAATTATTAGATGCCAAAGCCCGCAGCGGTTATCGCCTGGGC
GATGGTCAGGTGTACGACGTCATCTTACGCGATGGCTTAATGTGCGCGACCCACGGTTACCACATGGG
TATTACGGCCGAAAACGTGGCGAAAGAATACGGCATTACGCGCGAGATGCAGGATGAATTAGCACTG
CACTCTCAGCGCAAAGCAGCAGCCGCGATCGAGTCTGGTGCGTTTACGGCGGAAATCGTGCCAGTTAA
CGTGGTCACGCGCAAGAAGACGTTTCGTTTTACGCCAGGACGAGTTCCCGAAGGCAAACAGCACCGCG
GAGGCCTTAGGTGCCTTACGCCACGCCCTTGACAAAGCGGGCACGGTCACCGCCGGTAATGCGAGCGG
CATCAATGATGGTGCAGCGGCACTGGTCATCATGGAAGAGAGCGCCGCATTAGCAGCGGGTCTGACCC
CATTAGCGCGCATTAAATCTTATGCCAGCGGCGGCGTCCCACCAGCCCTGATGGGCATGGGTCCGGTC
CCAGCCACGCAAAAAGCCCTGCAATTAGCGGGCCTGCAACTGGCCGACATTGATCTGATCGAGGCGA
ACGAGGCGTTTGCAGCGCAGTTCCTGGCGGTGGGTAAGAATCTGGGCTTCGACAGCGAGAAAGTCAAT
GTGAACGGTGGCGCGAATTGCGTTAGGCCATCCGATTGGTGCAAGCGGCGCACGCATCTTAGTGACGTT
ACTGCACGCCATGCAGGCACGCGACAAGACCTTAGGCCTGGCGACCTTATGTATTGGTGGCGGTCAAG
GTATCGCCATGGTGATCGAACGCCTGAACTGAAGATCTAGGAGGAAAGCAAATGAACTGAGCACC
AAGCTGTGCTGGTGTGGCATCAAGGGTCGCCTGCGCCACAAAAGCAGCAACAGCTGCACAACACGA
ACCTGCAAATGACCGAGCTGAAAAAGCAGAAGACGGCCGAGCAAAAGACCCGCCCGCAGAACGTTGG
CATCAAGGGCATCCAGATTTATATCCCGACGCAGTGTGTCAACCAATCTGAGCTGGAGAAATTCGATG
GCGTCAGCCAGGGTAAGTACACCATCGGCCTGGGCCAGACCAACATGAGCTTCGTGAACGACCGTGA
GGACATCTATTCTATGAGCCTGACGGTGCTGTCTAAGCTGATCAAGAGCTACAACATCGACACGAATA
AGATCGGTCGTCTGGAGGTGGGTACGGAGACGCTGATTGACAAGAGCAAAAGCGTGAAGTCTGTCTT

AATGCAGCTGTTCCGGCGAGAACACGGATGTCGAGGGTATCGACACCCCTGAAUGGCTGTACGGCGGC
 ACCAACGCACTGTTCAATAGCCTGAACTGGATTGAGAGCAACGCCTGGGATGGCCGCGATGCGATCGT
 CGTGTGCGGCGATATCGCCATCTATGACAAGGGTGCGGCACGTCCGACCGGCGGTGCAGGCACCGTTG
 CGATGTGGATTGGCCCGGACGCACCAATTGTCTTCGATTCTGTCCGCGCGTCTTACATGGAGCACGCCT
 ACGACTTTTACAAGCCGGACTTCACGAGCGAATACCCGTACGTGGACGGCCACTTCTCTCTGACCTGCT
 ATGTGAAGGCGCTGGACCAGGTTTATAAGTCTTATAGCAAAAAGGCGATTTCTAAGGGCCTGGTCAGC
 GACCCGGCAGGCAGCGACGCCCTGAACGTGCTGAAGTATTTGACTACAACGTGTTCCATGTCCCGAC
 CTGCAAATTAGTGACCAAATCTTATGGCCGCTGTTATATAATGATTTCCGTGCCAACCCGCGAGCTGTT
 CCCGGAGGTTGACGCCGAGCTGGCGACGCGTGATTACGACGAGAGCCTGACCGACAAGAACATCGAG
 AAGACCTTCGTCAACGTGCGGAAGCCGTTCCACAAAGAGCGTGTGGCCCAAAGCCTGATCGTCCCGAC
 CAACACGGGCAACATGTATACCGCGTCTGTCTACGCGGCATTTCGCGAGCCTGCTGAATTACGTCGGTT
 CTGACGACCTGCAGGGCAAGCGCGTTGGCCTGTTTACGCTACGGTAGCGGCTTAGCGGCCAGGCTGTAT
 AGCTGCAAAATTGTGCGCGACGTCCAGCACATCATCAAGGAGCTGGACATCACCAACAAGCTGGCGA
 AGCGCATCACCGAGACGCCGAAAGATTACGAGGCAGCGATCGAGTTACGCGAGAATGCGCATCTGAA
 GAAGAACTTCAAGCCGCAAGGTAGCATCGAGCACCTGCAGAGCGGCGTCTACTACCTGACGAACATT
 GACGACAAGTTCCGCCGTTCTTATGACGTCAAAAAGTAACTAGTAGGAGGAAAACATCATGCAAAGTT
 TAGATAAGAATTTCCGACATTTATCTCGTCAACAAAAGTTACAACAATTGGTAGATAAGCAATGGTTA
 TCAGAAGATCAATTCGACATTTTATTGAATCATCCATTAATTGATGAGGAAGTAGCAAATAGTTTAATT
 GAAAATGTCATCGCGCAAGGTGCATTACCCGTTGGATTATTACCGAATATCATTGTGGACGATAAGGC
 ATATGTTGTACCTATGATGGTGGAAGAGCCTTCAGTTGTCGCTGCAGCTAGTTATGGTGCAAAGCTAGT
 GAATCAGACTGGCGGATTTAAAACGGTATCTTCTGAACGTATTATGATAGGTCAAATCGTCTTTGATG
 GCGTTGACGATACTGAAAAATTATCAGCAGACATTAAAGCTTTAGAAAAGCAAATTCATAAAATTGCG
 GATGAGGCATATCCTTCTATTAAAGCGCGTGGTGGTGGTTACCAACGTATAGCTATTGATACATTTCCCT
 GAGCAACAGTTACTATCTTTAAAAGTATTTGTTGATACGAAAGATGCTATGGGCGCTAATATGCTTAAT
 ACGATTTTAGAGGCCATAACTGCATTTTAAAAAATGAATCTCCACAAAGCGACATTTTAATGAGTATT
 TTATCCAATCATGCAACAGCGTCCGTTGTTAAAGTTCAAGGCGAAATTGACGTTAAAGATTTAGCAAG
 GGGCGAGAGAACTGGAGAAGAGGTTGCCAAACGAATGGAACGTGCTTCTGTATTGGCACAAGTTGAT
 ATTCATCGTGTGCAACACATAATAAAGGTGTTATGAATGGCATAACATGCCGTTGTTTTAGCAACAGG
 AAATGATACGCGTGGTGCAGAAGCAAGTGCATGCATACGCGAGTCGTGACGGACAGTATCGTGGT
 ATTGCAACATGGAGATACGATCAAAAACGTCAACGTTTAATTGGTACAATAGAAGTGCCTATGACATT
 GGCAATCGTTGGCGGTGGTACAAAAGTATTACCAATTGCTAAAGCTTCTTTAGAATTGCTAAATGTAG
 ATTCAGCACAGAATTAGGTCATGTAGTTGCTGCCGTTGGTTAGCACAGAAGTTGACGATGTGCGC
 GCGCTCGTTTCCGAAGGTATCCAGCAAGGCCATATGAGCTTGCAATATAAATCTTTAGCTATTGTTGTA
 GGTGCAAAAGGTGATGAAATTGCGCAAGTAGCTGAAGCATTGAAGCAAGAACCCCGTGCGAATACAC
 AAGTAGCTGAACGCATTTTACAAGAAATTAGACAACAATAG

SEQ ID NO: 42

atoB(opt):*mvaS*(opt):*mvaA* operon

ATGAAGAACTGTGTGATTGTTTCTGCGGTCCGCACGGCGATCGGCAGCTTTAACGGCTCTTTAGCGAG
 CACCTCTGCAATCGATCTGGGTGCGACGGTCATTAAGGCCGCCATTGAACGCGCCAAAATCGACAGCC
 AGCACGTTGATGAGGTGATCATGGGCAATGTGTTACAAGCCGGCCTGGGTCAAACCCAGCGCGTCAA
 GCACTGTTAAAATCTGGTCTGGCCGAGACCGTGTGTGGCTTCACCGTCAATAAGGTTTGCGGCTCTGGC

CTGAAGAGCGTGGCCCTGGCAGCACAAAGCGATTCAAGCCGGTCAGGCACAAAGCATCGTTGGGGTG
 GCATGGAGAACATGTCTCTGGCGCCGTACTIONATTAGATGCCAAAGCCCGCAGCGGTTATCGCCTGGGC
 GATGGTCAGGTGTACGACGTCATCTTACGCGATGGCTTAATGTGCGCGACCCACGGTTACCACATGGG
 TATTACGGCCGAAAACGTGGCGAAAGAATACGGCATTACGCGCGAGATGCAGGATGAATTAGCACTG
 CACTCTCAGCGCAAAGCAGCAGCCGCGATCGAGTCTGGTGCGTTTACGGCGGAAATCGTGCCAGTTAA
 CGTGGTCACGCGCAAGAAGACGTTTCGTTTTTCAGCCAGGACGAGTTCCCGAAGGCAAACAGCACCGCG
 GAGGCCTTAGGTGCCTTACGCCAGCCTTTGACAAAGCGGGCACGGTCACCGCCGGTAATGCGAGCGG
 CATCAATGATGGTGCAGCGGCACTGGTTCATCATGGAAGAGAGCGCCGCATTAGCAGCGGGTCTGACCC
 CATTAGCGCGCATTAAATCTTATGCCAGCGGCGGCGTCCCACCAGCCCTGATGGGCATGGGTCCGGTC
 CCAGCCACGCAAAAAGCCCTGCAATTAGCGGGCCTGCAACTGGCCGACATTGATCTGATCGAGGCGA
 ACGAGGCGTTTTGCAGCGCAGTTCCTGGCGGTGGGTAAGAATCTGGGCTTCGACAGCGAGAAAGTCAAT
 GTGAACGGTGGCGCGATTGCGTTAGGCCATCCGATTGGTGCAAGCGGCGCACGCATCTTAGTGACGTT
 ACTGCACGCCATGCAGGCACGCGACAAGACCTTAGGCCTGGCGACCTTATGTATTGGTGGCGGTCAAG
 GTATCGCCATGGTGATCGAACGCCTGAACTGAAGATCTAGGAGGAAAGCAAATGACAATAGGTATC
 GACAAAATAAACTTTTACGTTCCAAAGTACTATGTAGACATGGCTAAATTAGCAGAAGCACGCCAAGT
 AGACCCAAACAAATTTTTAATTGGAATTGGTCAAACCTGAAATGGCTGTTAGTCCTGTAAACCAAGACA
 TCGTTTCAATGGGCGCTAACGCTGCTAAGGACATTATAACAGACGAAGATAAAAAGAAAATTGGTATG
 GTAATTGTGGCAACTGAATCAGCAGTTGATGCTGCTAAAGCAGCCGCTGTTCAAATTCACAACTTATT
 AGGTATTCAACCTTTTGCACGTTGCTTTGAAATGAAAGAAGCTTGTTATGCTGCAACACCAGCAATTCA
 ATTAGCTAAAGATTATTTAGCAACTAGACCGAATGAAAAAGTATTAGTTATTGCTACAGATACAGCAC
 GTTATGGATTGAATTCAGGCGGCGAGCCAACACAAGGTGCTGGCGCAGTTGCGATGGTTATTGCACAT
 AATCCAAGCATTTTGGCATTAAATGAAGATGCTGTTGCTTACACTGAAGACGTTTATGATTTCTGGCGT
 CCAACTGGACATAAATATCCATTAGTTGATGGTGCATTATCTAAAGATGCTTATATCCGCTCATTCCAA
 CAAAGCTGGAATGAATACGCAAAACGTCAAGGTAAGTCGCTAGCTGACTTCGCATCTCTATGCTTCCA
 TGTTCCATTTACAAAAATGGGTAAAAAGGCATTAGAGTCAATCATTGATAACGCTGATGAAACAACCTC
 AAGAGCGTTTACGTTACAGGATATGAAGATGCTGTAGATTATAACCGTTATGTCCGTAATATTTATACTG
 GATCATTATATTTAAGCCTAATATCATTACTTGAAAATCGTGATTTACAAGCTGGTGAAACAATCGGTT
 TATTCAGTTATGGCTCAGGTTTCACTTGGTGAATTTTATAGTGCGACATTAGTTGAAGGCTACAAAGATC
 ATTTAGATCAAGCTGCACATAAAGCATTATTAATAAACCGTACTGAAGTATCTGTTGATGCATATGAA
 ACATTCTTCAAACGTTTTGATGACGTTGAATTTGACGAAGAACAAGATGCTGTTTATGAAGATCGTCAT
 ATTTTCTACTTATCAAATATTGAAAATAACGTTTCGCGAATATCACAGACCAGAGTAACTAGTAGGAGG
 AAAACATCATGCAAAGTTTAGATAAGAATTTCCGACATTTATCTCGTCAACAAAAGTTACAACAATTG
 GTAGATAAGCAATGGTTATCAGAAGATCAATTCGACATTTTATTGAATCATCCATTAATTGATGAGGA
 AGTAGCAAATAGTTTAATTGAAAATGTCATCGCGCAAGGTGCATTACCCGTTGGATTATTACCGAATA
 TCATTGTGGACGATAAGGCATATGTTGTACCTATGATGGTGGAAGAGCCTTCAGTTGTGCGCTGCAGCT
 AGTTATGGTGCAAAGCTAGTGAATCAGACTGGCGGATTTAAAACGGTATCTTCTGAACGTATTATGAT
 AGGTCAAATCGTCTTTGATGGCGTTGACGATACTGAAAAATTATCAGCAGACATTAAAGCTTTAGAAA
 AGCAAATTCATAAAATTGCGGATGAGGCATATCCTTCTATTAAAGCGCGTGGTGGTGGTTACCAACGT
 ATAGCTATTGATACATTTCTGAGCAACAGTTACTATCTTTAAAAGTATTTGTTGATACGAAAGATGCT
 ATGGGCGCTAATATGCTTAATACGATTTTAGAGGCCATAACTGCATTTTAAAAAATGAATCTCCACA
 AAGCGACATTTTAATGAGTATTTTATCCAATCATGCAACAGCGTCCGTTGTTAAAGTTCAAGGCGAAA
 TTGACGTTAAAGATTTAGCAAGGGGCGAGAGAACTGGAGAAGAGGTTGCCAAACGAATGGAACGTGC

TTCTGTATTGGCACAAGTTGATATTCATCGTGTGCAACACATAATAAAGGTGTTATGAATGGCATACA
TGCCGTTGTTTTAGCAACAGGAAATGATACGCGTGGTGCAGAAAGCAAGTGCGCATGCATACGCGAGTC
GTGACGGACAGTATCGTGGTATTGCAACATGGAGATACGATCAAAAACGTCAACGTTTAATTGGTACA
ATAGAAGTGCCTATGACATTGGCAATCGTTGGCGGTGGTACAAAAGTATTACCAATTGCTAAAGCTTC
TTTAGAATTGCTAAATGTAGATTCAGCACAGAATTAGGTCATGTAGTTGCTGCCGTTGGTTTAGCACA
GAACTTTGCAGCATGTCGCGCGCTCGTTTCCGAAGGTATCCAGCAAGGCCATATGAGCTTGCAATATA
AATCTTTAGCTATTGTTGTAGGTGCAAAAGGTGATGAAATTGCGCAAGTAGCTGAAGCATTGAAGCAA
GAACCCCGTGCGAATACACAAGTAGCTGAACGCATTTTACAAGAAATTAGACAACAATAG

SEQ ID NO: 43

pAM328 – ERG9-KANMX-MET3promoter-ERG9 (excluding vector backbone)

[illegible]

TGTGATCTCTAGCTTGGGTCTCTCTCTGTCTGCTAACAGTTGTGATATCGGCTGCCCTCATCTCGACCGGAT
GCAATGCCAATTGTAATAGCTTTCCCATGTTAATTATACTTTATTCTT

SEQ ID NO: 44

GAGTGAACCTGCTGCCTGGCGTGTCTGACTCAGTACATTTCATAGTGGATGGCGGCGTTAGTATC

SEQ ID NO: 45

CGTGTATACGTTTTCCGCTTCTGCTCTTCGTCTTTTCTCTTCTTCCGATATCACAACCTGTTACGA

SEQ ID NO: 46

GGTAAGACGGTTGGGTTTTATCTTTTGCAGTTGGTACTATTAAGAACAATCACAGGAAACAGCTATGA
CC

SEQ ID NO: 47

TTGCGTTTTGTACTTTGGTTTCGCTCAATTTTGCAGGTAGATAATCGAAAAGTTGTAAAACGACGGCCAG
T

SEQ ID NO: 48

pAM491 sequence (excluding vector backbone)

GTTTAAACTTGCTAAATTTCGAGTGAAACACAGGAAGACCAGAAAATCCTCATTTTCATCCATATTAACA
ATAATTTCAAATGTTTATTTGCATTATTTGAAACTAGGGAAGACAAGCAACGAAACGTTTTGAAAATTT
TGAGTATTTTCAATAAATTTGTAGAGGACTCAGATATTGAAAAAAAGCTACAGCAATTAATACTTGAT
AAGAAGAGTATTGAGAAGGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGT
CAAACGACGTTGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATGATAACAAACCGAA
GTTATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAGTGATGATATTTCATAAATAATGTAATT
CTATATATGTTAATTACCTTTTTTTCGAGGCATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGAA
GGTTAATGTGGCTGTGGTTTCAGGGTCCATACCCGGGAGTTATGACAATTACAACAACAGAATTCTTTC
TATATATGCACGAACCTTGTAATATGGAAGAAATTATGACGTACAAACTATAAAGTAAATATTTTACGT
AACACATGGTGCTGTTGTGCTTCTTTTTCAAGAGAATACCAATGACGTATGACTAAGTTTAGGATTTAA
TGCAGGTGACGGACCCATCTTTCAAACGATTTATATCAGTGGCGTCCAAATTGTTAGGTTTTGTTGGTT
CAGCAGGTTTTCTGTGTGGGTCATATGACTTTGAACCAAATGGCCGGCTGCTAGGGCAGCACATAAG
GATAATTCACCTGCCAAGACGGCACAGGCAACTATTCTTGCTAATTGACGTGCGTTGGTACCAGGAGC
GGTAGCATGTGGGCCTCTTACACCTAATAAGTCCAACATGGCACCTTGFGGTTCTAGAACAGTACCAC
CACCGATGGTACCTACTTCGATGGATGGCATGGATACGGAAATTCTCAAATCACCGTCCACTTCTTTCA
TCAATGTTATACAGTTGGAACCTTTCGACATTTTGTGCAGGATCTTGTCTAATGCCAAGAAAACAGCTG
TCACTAAATTAGCTGCATGTGCGTTAAATCCACCAACAGACCCAGCCATTGCAGATCCAACCAAATTC
TTAGCAATGTTCAACTCAACCAATGCGGAAACATCACTTTTAAACACTTTTCTGACAACATCACCAGGA
ATAGTAGCTTCTGCGACGACACTCTTACCACGACCTTCGATCCAGTTGATGGCAGCTGGTTTTTTGTGCG
GTACAGTAGTTACCAGAAACGGAGACAACCTCCATATCTTCCCAGCCATACTCTTCTACCATTTGCTTT
AATGAGTATTCGACACCCCTTAGAAATCATATTCATACCCATTGCGTCACCAGTAGTTGTTCTAAATCTC
ATGAAGAGTAAATCTCCTGCTAGACAAGTTTGAATATGTTGCAGACGTGCAAATCTTGATGTAGAGTT
AAAAGCTTTTTTAATTGCGTTTTGTCCCTCTTCTGAGTCTAACCATATCTTACAGGCACCAGATCTTTTC

AAAGTTG³⁷G³⁸AA³⁹CGGACTACTGGGCTCTTGTCATACCATCCTTAGTTAAAACAGTTGTTGCACCACC
GCCAGCATTGATTGCCTTACAGCCACGCATGGCAGAAGCTACCAAACAACCCTCTGTAGTTGCCATTG
GTATATGATAAGATGTACCATCGATAACCAAGGGGCTATAACACCAACGGGCAAAGGCATGTAACC
TATAACATTTTCACAACAAGCGCCAAATACGCGGTCTGTAGTCATAATTTTTATATGGTAAACGATCAG
ATGCTAATACAGGAGCTTCTGCCAAAATTGAAAGAGCCTTCCTACGTACCGCAACCGCTCTCGTAGTA
TCACCTAATTTTTCTCCAAAGCGTACAAAGGTAACCTACCGTGAATAACCAAGGCAGOGACCTCTTTG
TTCTTCAATTGTTTTGTATTTCCACTACTTAATAATGCTTCTAATCTTCTAAAGGACGTATTTTCTTATC
CAAGCTTTCAATATCGCGGGAATCATCTTCCTCACTAGATGATGAAGGTCCTGATGAGCTCGATTGCGC
AGATGATAAACTTTTGACTTTTCGATCCAGAAATGACTGTTTTATTGGTTAAAACTGGTGTAGAAGCCTT
TTGTACAGGAGCAGTAAAAGACTTCTTGGTGACTTCAGTCTTCACCAATTGGTCTGCAGCCATTATAGT
TTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTTTATGACATTTGAA
TAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTTTGCATTTAT
ATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCAGAAAATAAGGCTAAAAAACT
AATCGCATTATTATCCTATGGTTGTTAATTTGATTTCGTTGATTGAAGGTTTGTGGGGCCAGGTTACTG
CCAATTTTTCTCTTCATAACCATAAAAAGCTAGTATTGTAGAATCTTTATTGTTGGAGCAGTGCGGCG
CGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGT
CGGAGGGCTGTGCGCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTA
TACTGAAAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACAT
ATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTG
CGTCCATCCAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGAACTCTCAACTAACTTTG
TTGGTGTGGTATTAAAGGAAGACTTAGGCCGCAAAGCAACAACATTACACAATACAACTTGCAA
ATGACTGAACTAAAAAAACAAAAGACCGCTGAACAAAAAACAGACCTCAAAATGTCGGTATTAAAG
GTATCCAAATTTACATCCCAACTCAATGTGTCAACCAATCTGAGCTAGAGAAATTTGATGGCGTTTCTC
AAGGTAAATACACAATTGGTCTGGGCCAAACCAACATGTCTTTGTCAATGACAGAGAAGATATCTAC
TCGATGTCCCTAACTGTTTTGTCTAAGTTGATCAAGAGTTACAACATCGACACCAACAAAATTGGTAG
ATTAGAAGTCGGTACTGAACTCTGATTGACAAGTCCAAGTCTGTCAAGTCTGTCTTGATGCAATTGTT
TGGTGAAAACACTGACGTCGAAGGTATTGACACGCTTAATGCCTGTTACGGTGGTACCAACGCGTTGT
TCAACTCTTTGAACTGGATTGAATCTAACGCATGGGATGGTAGAGACGCCATTGTAGTTTGCGGTGAT
ATTGCCATCTACGATAAGGGTGCCGCAAGACCAACCGGTGGTGCCGGTACTGTTGCTATGTGGATCGG
TCCTGATGCTCCAATTGTATTTGACTCTGTAAGAGCTTCTTACATGGAACACGCCTACGATTTTTACAA
GCCAGATTTACCAGCGAATATCCTTACGTCGATGGTCATTTTTCATTAACCTTGTTACGTCAAGGCTCT
TGATCAAGTTTACAAGAGTTATTCCAAGAAGGCTATTTCTAAAGGGTTGGTTAGCGATCCCGCTGGTTC
GGATGCTTTGAACGTTTTGAAATATTTGACTACAACGTTTTCCATGTTCCAACCTGTAAATTGGTCAC
AAAATCATACGGTAGATTACTATATAACGATTTTCAGAGCCAATCCTCAATTGTTCCCAGAAGTTGACG
CCGAATTAGCTACTCGCGATTATGACGAATCTTTAACCGATAAGAACATTGAAAAAACTTTTGTTAAT
GTTGCTAAGCCATTCCACAAAGAGAGAGTTGCCCAATCTTTGATTGTTCCAACAAACACAGGTAACAT
GTACACCGCATCTGTTTATGCCGCCCTTTCATCTCTATTAAACTATGTTGGATCTGACGACTTACAAGG
CAAGCGTGTTGGTTTATTTTCTTACGGTTCCGGTTTAGCTGCATCTCTATATTCTTGCAAAATTGTTGGT
GACGTCCAACATATTATCAAGGAATTAGATATTACTAACAAATTAGCCAAGAGAATCACCGAACTCC
AAAGGATTACGAAGCTGCCATCGAATTGAGAGAAAATGCCCATTTGAAGAAGAACTTCAAACCTCAA
GGTTCCATTGAGCATTGCAAAGTGGTGTTTACTACTTGACCAACATCGATGACAAATTTAGAAGATCT
TACGATGTTAAAAATAATCTTCCCCCATCGATTGCATCTTGCTGAACCCCTTCATAAATGCTTTATTT

TTTTGGCAGCCTGCTTTTTTTAGCTCTCATTTAATAGAGTAGTTTTTTAATCTATAIACIAGGAAAACIC
 TTTATTTAATAACAATGATATATATATACCCGGGAAGCTTTTCAATTCATCTTTTTTTTTTTTGTCTTTT
 TTTTGATTCCGGTTTCTTTGAAATTTTTTTTGATTCCGTAATCTCCGAGCAGAAGGAAGAACGAAGGAAG
 GAGCACAGACTTAGATTGGTATATATACGCATATGTGGTGTGGAAGAAACATGAAATTGCCCAGTATT
 CTAAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTGAAAGCTACATAT
 AAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCACGAAAAGCA
 AACAACTTGTGTGCTTCATTGGATGTTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAG
 GTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTTCCATGGAGGGCAGAGTTA
 AGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTG
 GTAATACAGTCAAATTCGAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAAT
 GCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGG
 AACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAGGGGCTCCCTAGCTACTGGAGAATATACT
 AAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTATCGGCTTTATTGCTCAAAGAGACAT
 GGGTGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGA
 GACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATTATTGTT
 GGGTTTAAAC

SEQ ID NO: 49

pAM492 sequence (excluding vector backbone)

GTTTAAACTTGCTAAATTCGAGTGAAACACAGGAAGACCAGAAAATCCTCATTTTCATCCATATTAACA
 ATAATTTCAAATGTTTATTTGCATTATTTGAAACTAGGGAAGACAAGCAACGAAACGTTTTTGAAAATT
 TTGAGTATTTTCAATAAATTTGTAGAGGACTCAGATATTGAAAAAAGCTACAGCAATTAATACTTGA
 TAAGAAGAGTATTGAGAAGGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAG
 TCAAACGACGTTGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATGATAACAAACCGA
 AGTTATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAGTGATGATATTTCATAAATAATGTAAT
 TCTATATATGTTAATTACCTTTTTTTCGAGGCATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGA
 AGGTTAATGTGGCTGTGGTTTCAGGGTCCATACCCGGGTATATATATATCATTGTTATTAAATAAAGAG
 TTTTCTAGTATATAGATTAAAAAACTACTCTATTAAATGAGAGCTAAAAAAGCAGGCTGCCAAAAA
 AATAAAGCATTATGAAGGGGGTTCAGCAAGATGCAATCGATGGGGGAAGATTATTTTTTAACATCGT
 AAGATCTTCTAAATTTGTCATCGATGTTGGTCAAGTAGTAAACACCACTTTGCAAATGCTCAATGGAAC
 CTTGAGGTTTGAAGTTCTTCTTCAAATGGGCATTTTCTCTCAATTCGATGGCAGCTTCGTAATCCTTTGG
 AGTTTCGGTGATTCTCTTGCTAATTTGTTAGTAATATCTAATTCCTTGATAATATGTTGGACGTCACCA
 ACAATTTTGCAAGAATATAGAGATGCAGCTAAACCGGAACCGTAAGAAAATAAACCAACACGCTTGC
 CTTGTAAGTCGTCAGATCCAACATAGTTTAATAGAGATGCAAAGGCGGCATAAACAGATGCGGTGTAC
 ATGTTACCTGTGTTTGTGGAACAATCAAAGATTGGGCAACTCTCTCTTTGTGGAATGGCTTAGCAACA
 TTAACAAAAGTTTTTTCAATGTTCTTATCGGTTAAAGATTTCGTCATAATCGCGAGTAGCTAATTCGGCG
 TCAACTTCTGGGAACAATTGAGGATTGGCTCTGAAATCGTTATATAGTAATCTACCGTATGATTTTGTG
 ACCAATTTACAGGTTGGAACATGGAAAACGTTGTAGTCGAAATATTTCAAAACGTTCAAAGCATCCGA
 ACCAGCGGGATCGCTAACCAACCCTTTAGAAATAGCCTTCTTGGAATAACTCTTGTAACCTTGATCAA
 GAGCCTTGACGTAACAAGTTAATGAAAAATGACCATCGACGTAAGGATATTCGCTGGTGAAATCTGGC
 TTGTAAAAATCGTAGGCGTGTTCATGTAAGAAGCTCTTACAGAGTCAAATACAATTGGAGCATCAGG
 ACCGATCCACATAGCAACAGTACCGGCACCACCGGTTGGTCTTGCGGCACCCTTATCGTAGATGGCAA

TATCACCGCAA^{AAAA}CTACAATGGCGTCTCTACCATCCCATGCGTTAGATTCAATCCAGTTCAA^{AA}AGAGTTGA
ACAACGCGTTGGTACCACCGTAACAGGCATTAAGCGTGTCAATACCTTCGACGTCAGTGT^{TTT}TACCA
AACAATTGCATCAAGACAGACTTGACAGACTTGGACTTGTCAATCAGAGTTTCAGTACCGACTTCTAA
TCTACCAATTTTGT^TGGTGTGCGATGTTGTA^{ACT}CTTGATCAACTTAGACAAAACAGTTAGGGACATCGA
GTAGATATCTTCTCTGTCATTGACAAAAGACATGTTGGTTTGGCCCAGACCAATTGTGTATTTACCTTG
AGAAACGCCATCAAATTTCTCTAGCTCAGATTGGTTGACACATTGAGTTGGGATGTAAATTTGGATAC
CTTTAATACCGACATTTTGAGGTCTGGT^{TTTT}TGTTTCAGCGGTCTTTTGT^{TTTT}TTTAGTTTCAGTCATTTGC
AAGTTTGTATTGTGTAATTGT^TGGTGGCTTTTGGCGCCTAAGTCTTCCTTTAATACCACACCAACAAAGTT
TAGTTGAGAGTTTCATTTTATGTGATGATTGATTGATTGATTGTACAGTTTGT^{TTTT}TTCTTAATATCTATT
TCGATGACTTCTATATGATATTGCACTAACAAGAAGATATTATAATGCAATTGATACAAGACAAGGAG
TTATTTGCTTCTCTTTTATATGATTCTGACAATCCATATTGCGTTGGTAGTCTTTTGTCTGGAACGGTT
CAGCGGAAAAGACGCATCGCTCTTTTGTCTCTAGAAGAAATGCCAGCAAAAAGAAATCTCTTGACAGTG
ACTGACAGCAAAAATGTCTTTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGG
TGAAGTAATAATTAAATCATCCGTATAAACCTATACACATATATGAGGAAAAATAATACAAAAGTGTT
TTAAATACAGATACATACATGAACATATGCACGTATAGCGCCCAAATGTCGGTAATGGGATCGGCTTA
CTAATTATAAAATGCATCATAGAAATCGTTGAAGTTGACGCGAGCGACTCGAGATCCATAGGAGCAACT
CATGTCTGAAC^{TT}CAACGATTTCTATGATGCATTTTATAATTAGTAAGCCGATCCCATTACCGACATTT
GGGCGCTATACGTGCATATGTTCATGTATGTATCTGTATTTAAAACACTTTTGTATTATTTTCTCATA
TATGTGTATAGGTTTATACGGATGATTTAATTATTACTTCACCACCCTTTATTTTCAGGCTGATATCTTAG
CCTTGTTACTAGTTAGAAAAAGACATTTTGTCTGTCAGTCACTGTCAAGAGATTCTTTTGTCTGGCATT
CTTCTAGAAGCAAAAAGAGCGATGCGTCTTTTCCGCTGAACCGTTCCAGCAAAAAGACTACCAACGC
AATATGGATTGTCAGAATCATATAAAAGAGAAGCAAATAACTCCTTGTCTTGTATCAATTGCATTATA
ATATCTTCTTGTAGTGCAATATCATATAGAAGTCATCGAAATAGATATTAAGAAAAACAACTGTAC
AATCAATCAATCAATCATCACATAAAATGGCTGCAGACCAATTGGTGAAGACTGAAGTCACCAAGAA
GTC^{TTTT}TTACTGCTCCTGTACAAAAGGCTTCTACACCAGTTTAAACCAATAAAACAGTCATTTCTGGATC
GAAAGTCAAAAGTTTATCATCTGCGCAATCGAGCTCATCAGGACCTTCATCATCTAGTGAGGAAGATG
ATTCCCGGATATTGAAAGCTTGGATAAGAAAATACGTCCTTTAGAAGAATTAGAAGCATTATTAAGT
AGTGGAATAACAAAACAATTGAAGAACAAGAGGTCGCTGCCTTGGTTATTACCGGTAAGTTACCTTT
GTACGCTTTGGAGAAAAAATTAGGTGATACTACGAGAGCGGTTGCGGTACGTAGGAAGGCTCTTCAA
TTTTGGCAGAAGCTCCTGTATTAGCATCTGATCGTTTACCATATAAAAATTATGACTACGACCGCGTAT
TTGGCGCTTGTGTGAAAATGTTATAGGTTACATGCCTTTGCCCCTTGGTGTATAGGCCCTTGGTTAT
CGATGGTACATCTTATCATATACCAATGGCAACTACAGAGGGTTGTTTGGTAGCTTCTGCCATGCGTGG
CTGTAAGGCAATCAATGCTGGCGGTGGTGCAACAACTGTTTTAACTAAGGATGGTATGACAAGAGGCC
CAGTAGTCCGTTTCCCAACTTTGAAAAGATCTGGTGCCTGTAAGATATGGTTAGACTCAGAAGAGGGA
CAAAACGCAATTAAAAAAGCTTTTAACTCTACATCAAGATTTGCACGTCTGCAACATATTCAAAC^{TT}TG
CTAGCAGGAGATTTACTCTTCATGAGATTTAGAACA^{ACT}ACTGGTGACGCAATGGGTATGAATATGAT
TTCTAAGGGTGTGGAATACTCATTAAGCAAATGGTAGAAGAGTATGGCTGGGAAGATATGGAGGTTG
TCTCCGTTTCTGGTAACTACTGTAECGACAAAAAACAGCTGCCATCAACTGGATCGAAGGTCGTGGT
AAGAGTGTGTCGCGAGAAGCTACTATTCCTGGTGTGTTGTGCAAAAAGTGTTAAAAAAGTGATGTTTC
CGCATTTGGTTGAGTTGAACATTGCTAAGAATTTGGTTGGATCTGCAATGGCTGGGTCTGTTGGTGGATT
TAACGCACATGCAGCTAATTTAGTGACAGCTGTTTCTTGGCATTAGGACAAGATCCTGCACAAAATG
TCGAAAGTTCCA^{ACT}GTATAACATTGATGAAAGAAGTGGACGGTGATTTGAGAATTTCCGTATCCATG

CCATCCATCGAAGT~~AGG~~TACCATCGGTGGTGGTAC~~TT~~CTAGAACCAAGGTGCCATGTTGGACTT
ATTAGGTGTAAGAGGCCACATGCTACCGCTCCTGGTACCAACGCACGTCAATTAGCAAGAATAGTTG
CCTGTGCCGTCTTGGCAGGTGAATTATCCTTATGTGCTGCCCTAGCAGCCGGCCATTTGGTTCAAAGTC
ATATGACCCACAACAGGAAACCTGCTGAACCAACAAAACCTAACAATTTGGACGCCACTGATATAAAT
CGTTTGAAAGATGGGTCCGTACCTGCATTAAATCCTAAACTTAGTCATACGTCATTGGTATTCTCTTG
AAAAAGAAGCACAAACAGCACCATGTGTTACGTAAAATATTTACTTTATAGTTTGTACGTCATAATTTCT
TCCATATTACAAGTTCGTGCATATATAGAAAGAATTCTGTTGTTGTAATTGTCATAACTCCCGGGAAGC
TTTTCAATTCATCTTTTTTTTTTTTGTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTTTTTGATTCCGTA
ATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATGTG
GTGTTGAAGAAACATGAAATTGCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAA
ACGAAGATAAATCATGTGCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTG
CCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTTCGTACCACC
AAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAATTTGTTTACTAAAAACACATGTGGATAT
CTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTT
ACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTAT
ACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGG
TTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCCTTTTGATGTTAGCAGAATTGTCAT
GCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAAGA
TTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTAT
GACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGAT
GTGGTCTCTACAGGATCTGACATTATTATTGTTGGGTTTAAAC

SEQ ID NO: 50

pAM489 sequence (excluding vector backbone)

GTTTAAACTACTATTAGCTGAATTGCCACTGCTATCGTTGTTAGTGGCGTTAGTGCTTGCATTCAAAGA
CATGGAGGGCGTTATTACGCCGGAGCTCCTCGACAGCAGATCTGATGACTGGTCAATATATTTTTGCAT
TGAGGCTCTGTTTGAATTATATTTTGAGATGACCCATCTAATGTACTGGTATCACCAGATTTTCATGTC
GTTTTTTAAAGCGGCTGCTTGAGTCTTAGCAATAGCGTCACCATCTGGTGAATCCTTTGAAGGAACCAC
TGACGAAGGTTTGGACAGTGACGAAGAGGATCTTTCTGCTTTGAATTAGTCGCGCTGGGAGCAGATG
ACGAGTTGGTGGAGCTGGGGGCAGGATTGCTGGCCGTCGTGGGTCTGAATGGGTCTTTGGCTGGTCC
ATCTCTATTCTGAAAACGGAAGAGGAGTAGGGAATATTACTGGCTGAAAATAAGTCTTGAATGAACGT
ATACGCGTATATTTCTACCAATCTCTCAACACTGAGTAATGGTAGTTATAAGAAAGAGACCGAGTTAG
GGACAGTTAGAGGCGGTGGAGATATTCCTTATGGCATGTCTGGCGATGATAAAACTTTTCAAACGGCA
GCCCCGATCTAAAAGAGCTGACACCCGGGAGTTATGACAATTACAACAACAGAATTCTTTCTATATAT
GCACGAACTTGTAATATGGAAGAAATTATGACGTACAACTATAAAGTAAATATTTTACGTAACACAT
GGTGCTGTTGTGCTTCTTTTTCAAGAGAATACCAATGACGTATGACTAAGTTTAGGATTTAATGCAGGT
GACGGACCCATCTTTCAAACGATTTATATCAGTGGCGTCCAAATTGTTAGGTTTTGTTGGTTCAGCAGG
TTTCTGTTGTGGGTCATATGACTTTGAACCAAATGGCCGGCTGCTAGGGCAGCACATAAGGATAATT
CACCTGCCAAGACGGCACAGGCAACTATTCTTGCTAATTGACGTGCGTTGGTACCAGGAGCGGTAGCA
TGTGGGCCTCTTACACCTAATAAGTCCAACATGGCACCTTGTGGTTCTAGAACAGTACCACCACCGAT
GGTACCTACTTCGATGGATGGCATGGATACGGAAATTCTCAAATCACCGTCCACTTCTTTTCATCAATGT
TATACAGTTGGAACCTTCGACATTTTGTGCAGGATCTTGTCTAATGCCAAGAAAACAGCTGTCACTAA

ATTAGCTG³C⁴ATGTG⁵CGTTAAATCCACCAACAGACCCAGCCATTGCAGATCCAACCAAATTCCTTAGCAA
 TGTTCAACTCAACCAATGCGGAAACATCACTTTTAAACACTTTTCTGACAACATCACCAGGAATAGTAG
 CTTCTGCGACGACACTCTTACCACGACCTTCGATCCAGTTGATGGCAGCTGGTTTTTTGTCGGTACAGT
 AGTTACCAGAAACGGAGACAACCTCCATATCTTCCCAGCCATACTCTTCTACCATTGTCTTAATGAGT
 ATTCGACACCCTTAGAAATCATATTCATACCCATTGCGTCACCAGTAGTTGTTCTAAATCTCATGAAGA
 GTAAATCTCCTGCTAGACAAGTTTGAATATGTTGCAGACGTGCAAATCTTGATGTAGAGTTAAAGCT
 TTTTAAATTGCGTTTTGTCCCTCTTCTGAGTCTAACCATATCTTACAGGCACCAGATCTTTTCAAAGTTG
 GGAAACGGACTACTGGGCCTCTTGTCTATACCATCCTTAGTTAAAAACAGTTGTTGCACCACCGCCAGCA
 TTGATTGCCTTACAGCCACGCATGGCAGAAGCTACCAAACAACCTCTGTAGTTGCCATTGGTATATG
 ATAAGATGTACCATCGATAACCAAGGGGCTATAACACCAACGGGCAAAGGCATGTAACCTATAACA
 TTTTCACAACAAGCGCCAAATACGCGGTCTGTAGTCATAATTTTTATATGGTAAACGATCAGATGCTAAT
 ACAGGAGCTTCTGCCAAAATTGAAAGAGCCTTCCTACGTACCGCAACCGCTCTCGTAGTATCACCTAA
 TTTTTCTCCAAAGCGTACAAAGGTAACCTACCGTGAATAACCAAGGCAGCGACCTCTTTGTTCTTCAA
 TTGTTTTGTATTTCCACTACTTAATAATGCTTCTAATTCTTCTAAAGGACGTATTTCTTATCCAAGCTTT
 CAATATCGCGGGAATCATCTTCCTCACTAGATGATGAAGGTCCTGATGAGCTCGATTGCGCAGATGAT
 AAACTTTTGACTTTTCGATCCAGAAATGACTGTTTTIATTGGTTAAAACTGGTGTAGAAGCCTTTTGTACA
 GGAGCAGTAAAAGACTTCTTGGTGACTTCAGTCTTCACCAATTGGTCTGCAGCCATTATAGTTTTTTCT
 CCTTGACGTAAAGTATAGAGGTATATTAACAATTTTTTGTGTGATACTTTTATGACATTTGAATAAGAA
 GTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTTTGCATTTATATATCTG
 TTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCAGAAATAAGGCTAAAAAACTAATCGCA
 TTATTATCCTATGGTTGTAAATTTGATTCTGTTGATTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTT
 TCCTCTTCATAACCATAAAAGCTAGTATTGTAGAATCTTTATTGTTTCGGAGCAGTGCGGCGCGAGGCAC
 ATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGC
 TGTCGCCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGAAA
 GTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACATATAAGTAA
 GATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTGCGTCCATC
 CAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGGCTTCAGAAAAAGAAATTAGGAGAGAG
 AGATTCTTGAACGTTTTCCCTAAATTAGTAGAGGAATTGAACGCATCGTTTTGGCTTACGGTATGCCT
 AAGGAAGCATGTGACTGGTATGCCCCTCATTGAACTACAACACTCCAGGCGGTAAGCTAAATAGAG
 GTTTGTCCGTTGTGGACACGTATGCTATTCTCTCCAACAAGACCGTTGAACAATTGGGGCAAGAAGAA
 TACGAAAAGGTTGCCATTCTAGGTTGGTGCATTGAGTTGTTGCAGGCTTACTTCTTGGTCCCGATGAT
 ATGATGGACAAGTCCATTACCAGAAGAGGCCAACCATGTTGGTACAAGGTTCCCTGAAGTTGGGGAAAT
 TGCCATCAATGACGCATTCATGTTAGAGGCTGCTATCTACAAGCTTTTGAAATCTCACTTCAGAAACGA
 AAAATACTACATAGATATCACCGAATTGTTCCATGAGGTCACCTTCCAAACCGAATTGGGCCAATTGA
 TGGACTTAATCACTGCACCTGAAGACAAAGTCGACTTGAGTAAAGTTCTCCCTAAAGAAGCACTCCTTC
 ATAGTTACTTTCAAGACTGCTTACTATTCTTTCTACTTGCTGTCGATTGGCCATGTACGTTGCCGGTA
 TCACGGATGAAAAGGATTTGAAACAAGCCAGAGATGTCTTGATTCCATTGGGTGAATACTTCCAAATT
 CAAGATGACTACTTAGACTGCTTCGGTACCCAGAACAGATCGGTAAGATCGGTACAGATATCCAAGA
 TAACAAATGTTCTTGGGTAAATCAACAAGGCATTGGAACCTGCTTCCGCAGAACAAAGAAAGACTTTAG
 ACGAAAATTACGGTAAGAAGGACTCAGTCGCAGAAGCCAAATGCAAAAAGATTTTCAATGACTTGAA
 AATTGAACAGCTATACCACGAATATGAAGAGTCTATTGCCAAGGATTTGAAGGCCAAAATTTCTCAGG
 TCGATGAGTCTCGTGGCTTCAAAGCTGATGTCTTAACCTGCGTTCTTGAACAAAGTTTACAAGAGAAGC

AAATAGAACTAAACGCTAAATCGATAAAAACATTAGATTTCAAACTAGATAAGGACCAIGTATAAAGAACTA
TATACTTCCAATATAATATAGTATAAGCTTTAAGATAGTATCTCTCGATCTACCGTTCCACGTGACTAG
TCCAAGGATTTTTTTTAAACCCGGGATATATGTGTACTTTGCAGTTATGACGCCAGATGGCAGTAGTGGA
AGATATTCTTTATTGAAAAATAGCTTGTACCTTACGTACAATCTTGATCCGGAGCTTTTCTTTTTTGC
CGATTAAGAATTCTGGTCGAAAAAAGAAAAGGAGAGGGCCAAGAGGGAGGGCATTGGTGACTATTGAG
CACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTATGTCTGTTATTAATTCACAGGTA
GTTCTGGTCCATTGGTGAAAGTTTGC GGCTTGCAGAGCACAGAGGCCCGCAGAATGTGCTCTAGATTCC
GATGCTGACTTGCTGGGTATTATATGTGTGCCCAATAGAAAGAGAACAATTGACCCGGTTATTGCAAG
GAAAATTTCAAGTCTTGTAAGCATATAAAAATAGTTCAGGCACTCCGAAATACTTGGTTGGCGTGT
TTCGTAATCAACCTAAGGAGGATGTTTTGGCTCTGGTCAATGATTACGGCATTGATATCGTCCAACCTGC
ATGGAGATGAGTCGTGGCAAGAATACCAAGAGTTCCTCGTTTTGCCAGTTATTAAGAACTCGTATTT
CCAAAAGACTGCAACATACTACTCAGTGCAGCTTCACAGAAACCTCATTGTTTTATTCCCTTGTTTGAT
TCAGAAGCAGGTGGGACAGGTGAACTTTTGGATTGGAACTCGATTTCTGACTGGGTTGGAAGGCAAGA
GAGCCCCGAAAGCTTACATTTTATGTTAGCTGGTGGACTGACGCCGTTTAAAC

SEQ ID NO: 51

pAM497 sequence (excluding vector backbone)

GTTTAAACTTTTCCAATAGGTGGTTAGCAATCGTCTTACTTTCTAACTTTTCTTACCTTTTACATTTTCA
CAATATATATATATATATTTTCAAGGATATACCATTTCTAATGTCTGCCCTAAGAAGATCGTCGTTTTGC
CAGGTGACCACGTTGGTCAAGAAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTT
CGTTCCAATGTCAAGTTTCGATTTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTT
CCACTTCCAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGTGGGTGG
TCCTAAATGGGGTACCGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATCCGTAAAGAACTTCAAT
TGACGCCAACTTAAGACCATGTAACCTTGCATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCAC
AATTTGCTAAAGGTACTGACTTCGTTGTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGA
AAGGAAGACGTTTAGCTTGCTCGTCCCCGCCGGGTACCCGGCCAGCGACATGGAGGCCCAGAATAC
CCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTGCCCCGTACATTTAGCCCAT
ACATCCCCATGTATAATCATTTGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACG
GCTCCTCGCTGCAGACCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCC
GCGCCCCTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTTCATATACTTCCTTTT
AAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGCAGAACCCAGCCCCAA
AAAAAGCAAAAACAACTGTTTCAGGAGCGCAAGGCGTTTATCTCCCGTATCACTAATGAAACTAAAAT
TCAAATCGCTATTTTCGCTGAATGGTGGTTATATTCAAATAAAAGATTTCGATTCTTCCTGCAAAGAAGGA
TGACGATGTAGCTTCCCAAGCTACTCAGTCACAGGTCATCGATATTCACACAGGTGTTGGCTTTTTGGA
TCATATGATCCATGCGTTGGCAAAACACTCTGGTTGGTCTCTTATTGTTGAATGTATTGGTGACCTGCA
CATTGACGATCACCATACTACCGAAGATTGCGGTATCGCATTAGGGCAAGCGTTCAAAGAAGCAATGG
GTGCTGTCCGTGGTGTA AAAAGATTCTGGTACTGGGTTTCGCACCATTTGGATGAGGCGCTATCACGTGCC
GTAGTCGATTTATCTAGTAGACCATTTGCTGTAATCGACCTTGGATTGAAGAGAGAGATGATTGGTGA
TTTATCCACTGAAATGATTCCACACTTTTTTGGAAGTTTCGCGGAGGCGGCCAGAATTACTTTGCATGT
TGATTGTCTGAGAGGTTTCAACGATCACCACAGAAGTGAGAGTGC GTTCAAGGCTTTGGCTGTGCCA
TAAGAGAAGCTATTTCTAGCAATGGCACCAATGACGTTCCCTCAACCAAAGGTGTTTTGATGTGAAGT
ACTGACAATAAAAAGATTCTTGTTTTCAAGAACTTGTCAATTTGTATAGTTTTTTTATATTGTAGTTGTTT

TATTTTAAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCCTCGACATCATCTGCCAGATGCGAAGTT
AAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTGCTGTGCGATT
GATACTAACGCCGCCATCCACCCGGGTTTCTCATTCAAGTGGTAACTGCTGTTAAAATTAAGATATTTA
TAAATTGAAGCTTGGTCGTTCCGACCAATACCGTAGGGAAACGTAAATTAGCTATTGTAAAAAAGGA
AAAGAAAAGAAAAGAAAAATGTTACATATCGAATTGATCTTATTCCCTTGGTAGACCAGTCTTTGCGT
CAATCAAAGATTGTTTTGTTTCTTGTGGGCCTGAACCGACTTGAGTTAAAATCACTCTGGCAACATCCT
TTTGCAACTCAAGATCCAATTCACGTGCAGTAAAGTTAGATGATTCAAATTGATGGTTGAAAGCCTCA
AGCTGCTCAGTAGTAAATTTCTTGTCCCATCCAGGAACAGAGCCAAACAATTTATAGATAAATGCAAA
GAGTTTCGACTCATTTTCAGCTAAGTAGTACAACACAGCATTTGGACCTGCATCAAACGTGTATGCAA
CGATTGTTTTCTCCGTAAAACTGATTAATGGTGTGGCACCAACTGATGATACGCTTGGAAGTGTCAATTCA
TGTAAGAATATTGGAGGGAAAGAGTCCAAACATGTGGCATGGAAAGAGTTGGAATCCATCATTGTTTTCC
TTTGCAAAGGTGGCGAAATCTTTTTCAACAATGGCTTTACGCATGACTTCAAATCTCTTTGGTACGACA
TGTTCAATTCTTTCTTTAAATAGTTCCGAGGTTGCCACGGTCAATTGCATACCCTGAGTGGAACCTCACA
TCCTTTTAAATATCGCTGACAACCTAGGACACAAGCTTTCATCTGAGGCCAGTCAGAGCTGTCTGCGATT
TGTAAGTCCATGGAATCATGACCATCTTCAGCTTTTCCCATTTCCCAGGCCACGTATCCGCCAAACAAC
GATCTACAAGCTGAACCAGACCCCTTCTTGTCTATTCTAGATATTTCTGAAGTTGACTGTGGTAATTGG
TATAACTTAGCAATTGCAGAGACCAATGCAGCAAAGCCAGCAGCGGAGGAAGCTAAACCAGCTGCTG
TAGGAAAGTTATTTTCGGAGACAATGTGGAGTTTCCATTGAGATAATGTGGGCAATGAGGCGTCCTTC
GATTCCATTTCCCTTTCTTAATTGGCGTAGGTGCGCGCAGACAATTTTGAGTTCTTTCAATTGTGCGATGCTGT
GTGGTTCTCCATTAAACCACAAAGTGTGCGGTTCAAACCTCAGGTGCAGTAGCCGCAGAGGTCAACGTT
CTGAGGTGATCTTGCGATAAAGTCACTGATATGGACGAATTGGTGGGCAGATTCAAACCTCGTGTCCCTT
TTCCCCCAATACTTAAGGGTTGCGATGTTGACGGGTGCGGTAACGGATGCTGTGTAAACGGTCATTAT
AGTTTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTTTATGACATTT
GAATAAGAAGTAATACAAACCGAAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTTTGCATT
TATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCAGAAATAAGGCTAAAAA
ACTAATCGCATTATTATCCTATGGTTGTTAATTTGATTGCTTGATTTGAAGGTTTGTGGGGCCAGGTTA
CTGCCAATTTTTCTCCTTTCATAACCATAAAAGCTAGTATTGTAGAATCTTTATTGTTTCGGAGCAGTGCG
GCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTC
CGTCGGAGGGCTGTGCCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGC
GTATTACTGAAAGTTCCAAAGAGAAGGTTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAA
CATATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAACTTCT
TTGCGTCCATCCAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGTCAGAGTTGAGAGCCTT
CAGTGCCCCAGGGAAAGCGTTACTAGCTGGTGGATATTTAGTTTTAGATCCGAAATATGAAGCATTG
TAGTCGGATTATCGGCAAGAATGCATGCTGTAGCCCATCCTTACGGTTCATTGCAAGAGTCTGATAAG
TTTGAAGTGCGTGTGAAAAGTAAACAATTTAAAGATGGGGAGTGGCTGTACCATATAAGTCCTAAAAC
TGGCTTCATTCTGTTTCGATAGGCGGATCTAAGAACCCTTTTCATTGAAAAAGTTATCGCTAACGTATT
TAGCTACTTTAAGCCTAACATGGACGACTACTGCAATAGAACTTGTTCGTTATTGATATTTTCTCTGA
TGATGCCTACCATTCTCAGGAGGACAGCGTTACCGAACATCGTGGGAACAGAAGATTGAGTTTTCAATT
CGCACAGAATTGAAGAAGTTCCCAAAACAGGGCTGGGCTCCTCGGCAGGTTTAGTCACAGTTTTAACT
ACAGCTTTGGCCTCCTTTTTTGTATCGGACCTGGAAAATAATGTAGACAAATATAGAGAAGTTATTCAT
AATTTATCACAAGTTGCTCATTGTCAAGCTCAGGGTAAAATTGGAAGCGGGTTTGATGTAGCGGCGGC
AGCATATGGATCTATCAGATATAGAAGATTCCCACCCGCATTAATCTCTAATTTGCCAGATATTGGAA

GTGCTACTTACGGCAGTAAACTGGGUCATTTGGTTAAATGAAGAAGACTGGAAATATACCGATTAAAAGT
AACCATTTACCTTCGGGATTAACCTTATGGATGGGCGATATTAAGAATGGTTCAGAAACAGTAAAACCT
GGTCCAGAAGGTAAAAAATTGGTATGATTTCGCATATGCCGGAAAGCTTGAAAATATATACAGAACTCG
ATCATGCAAATTCTAGATTTATGGATGGACTATCTAAACTAGATCGCTTACACGAGACTCATGACGATT
ACAGCGATCAGATATTTGAGTCTCTTGAGAGGAATGACTGTACCTGTCAAAAAGTATCCTGAGATCACA
GAAGTTAGAGATGCAGTTGCCACAATTAGACGTTCCCTTTAGAAAAATAACTAAAGAATCTGGTGCCGA
TATCGAACCTCCCGTACAAACTAGCTTATTGGATGATTGCCAGACCTTAAAAGGAGTTCCTTACTTGCTT
AATACCTGGTGCTGGTGGTTATGACGCCATTGCAGTGATTGCTAAGCAAGATGTTGATCTTAGGGCTC
AAACCGCTGATGACAAAAGATTTTCTAAGGTTCAATGGCTGGATGTAACCTCAGGCTGACTGGGGTGTT
AGGAAAGAAAAAGATCCGGAAACTTATCTTGATAAAATAACTTAAGGTAGATAATAGTGGTCCATGTG
ACATCTTTATAAATGTGAAGTTTGAAGTGACCGCGCTTAACATCTAACCATTCTCTCCGATAGTACT
TGAAATTGTTCCCTTTTCGGCGGCATGATAAAATTCTTTTAATGGGTACAAGCTACCCGGGAAAGATTCTC
TTTTTTTATGATATTTGTACATAAACTTTATAAATGAAATTCATAATAGAAACGACACGAAATTACAAA
ATGGAATATGTTTCATAGGGTAGACGAAACTATATACGCAATCTACATACATTTATCAAGAAGGAGAAA
AAGGAGGATGTAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTCAACGTGATAAGGAAAAAGA
ATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCACACAAAAAGTTAGGTGTAACAGAAAA
TCATGAAACTATGATTCCTAATTTATATATTGGAGGATTTTCTCTAAAAAAAAAAAAATACAACAAAT
AAAAAACACTCAATGACCTGACCATTTGATGGAGTTTAAGTCAATACCTTCTTGAACCATTTCCCATAA
TGGTGAAAGTTCCCTCAAGAATTTTACTCTGTGAGAAACGGCCTTAACGACGTAGTCGACCTCCTCTTC
AGTACTAAATCTACCAATACCAATCTGATGGAAGAATGGGCTAATGCATCATCCTTACCCAGCGCAT
GTAAACATAAGAAGGTTCTAGGGAAGCAGATGTACAGGCTGAACCCGAGGATAATGCGATATCCCT
TAGTGCCATCAATAAAGATTCTCCTTCCACGTAGGCGAAAGAAACGTTAACACGTTTAAAC

SEQ ID NO: 52

pAM493 sequence (excluding vector backbone)

GTTTAAACTACTCAGTATATTAAGTTTCGAATTGAAGGGCGAACTCTTATTCGAAGTCGGAGTCACCAC
AACACTTCCGCCCATACTCTCCGAATCCCTCGTTTCCTAAAGTAAGTTTACTTCCACTTGTAGGCCTATTA
TTAATGATATCTGAATAATCCTCTATTAGGGTTGGATCATTCAAGTAGCGCGTGCATTGAAAGGAGTCC
ATGCCCCGACGTGACGTGATTAGCGAAGGCGCGTAACCATTTGTCATGTCTAGCAGCTATAGAACTAAC
CTCCTTGACACCACTTGCGGAAGTCTCATCAACATGCTCTTCCTTATTACTCATTCTCTTACCAAGCAG
AGAATGTTATCTAAAAACTACGTGTATTTACCTCTTTCTCGACTTGAACACGTCCAACCTCCTTAAGTA
CTACCACAGCCAGGAAAGAATGGATCCAGTTCTACACGATAGCAAAGCAGAAAAACACAACCAGCGTA
CCCCTGTAGAAGCTTCTTTGTTTACAGCACTTGATCCATGTAGCCATACTCGAAATTTCAACTCATCTG
AACTTTTTCCTGAAGGTTGAAAAAGAATGCCATAAGGGTCACCCGAAGCTTATTCACGCCCCGGGAGTT
ATGACAATTACAACAACAGAATTTCTTTCTATATATGCACGAACCTTGTAATATGGAAGAAATTATGACG
TACAAACTATAAAGTAAATATTTTACGTAAACACATGGTGCTGTTGTGCTTCTTTTTCAAGAGAATACCA
ATGACGTATGACTAAGTTTAGGATTTAATGCAGGTGACGGACCCATCTTTCAAACGATTTATATCAGTG
GCGTCCAAATTGTTAGGTTTTGTTGGTTCAGCAGGTTTCCTGTTGTGGGTTCATATGACTTTGAACCAA
TGGCCGGCTGCTAGGGCAGCACATAAGGATAATTCACCTGCCAAGACGGCACAGGCAACTATTCTTGC
TAATTGACGTGCGTTGGTACCAGGAGCGGTAGCATGTGGGCCTCTTACACCTAATAAGTCCAACATGG
CACCTTGTGGTTCTAGAACAGTACCACCACCGATGGTACCTACTTCGATGGATGGCATGGATACGGAA
ATTCTCAAATCACCGTCCACTTCTTTCATCAATGTTATACAGTTGGAACCTTCGACATTTTGTGCAGGAT

CTTGTCC¹¹²TAATGCCAAGAAAACAGCTGTCACTAAATTAGCTGCATGTGCGTTAAATCCACCAACAGAC
 CCAGCCATTGCAGATCCAACCAAAT
 TCTTAGCAATGTTCAACTCAACCAATGCGGAAACATCACTTTTTTAACACTTTTCTGACAACATCACCAG
 GAATAGTAGCTTCTGCGACGACACTCTTACCACGACCTTCGATCCAGTTGATGGCAGCTGGTTTTTTGT
 CGGTACAGTAGTTACCAGAAACGGAGACAACTCCATATCTTCCCAGCCATACTCTTCTACCATTTGCT
 TTAATGAGTATTCGACACCCTTAGAAATCATATTCATACCCATTGCGTCACCAGTAGTTGTTCTAAATC
 TCATGAAGAGTAAATCTCCTGCTAGACAAAGTTTGAATATGTTGCAGACGTGCAAATCTTGATGTAGAG
 TAAAAAGCTTTTTTAATTGCGTTTTGTCCCTCTTCTGAGTCTAACCATATCTTACAGGCACCAGATCTTT
 TCAAAGTTGGGAAACGGACTACTGGGCCTCTTGTGCATACCATCCTTAGTTAAAACAGTTGTTGCACCAC
 CGCCAGCATTGATTGCCTTACAGCCACGCATGGCAGAAGCTACCAAACAACCCCTCTGTAGTTGCCATT
 GGTATATGATAAGATGTACCATCGATAACCAAGGGGCTATAACACCAACGGGGCAAAGGCATGTAAC
 CTATAACATTTTCACAACAAGCGCCAAATACGCGGTCTGAGTCATAATTTTTATATGGTAAACGATCA
 GATGCTAATACAGGAGCTTCTGCCAAAATTGAAAGAGCCTTCCTACGTACCGCAACCGCTCTCGTAGT
 ATCACCTAATTTTTTCTCCAAAGCGTACAAAGGTAACCTACCGTGAATAACCAAGGCAGCGACCTCTTT
 GTTCTTCAATTGTTTTGTATTTCCACTACTTAATAATGCTTCTAATTCTTCTAAAGGACGTATTTCTTAT
 CCAAGCTTTCAATATCGCGGGAATCATCTTCTCACTAGATGATGAAGGTCCTGATGAGCTCGATTGCG
 CAGATGATAAACTTTTGACTTTTCGATCCAGAAATGACTGTTTTATTGGTTAAAACCTGGTGTAGAAGCCT
 TTTGTACAGGAGCAGTAAAAGACTTCTTGGTGACTTCAGTCTTCACCAATTGGTCTGCAGCCATTATAG
 TTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTTTATGACATTTGA
 ATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGTGTTATGCAGCTTTTGCATTTA
 TATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCCAGAAATAAGGCTAAAAAAC
 TAATCGCATTATTATCCTATGGTTGTTAATTTGATTCTGTTGATTTGAAGGTTTGTGGGGCCAGGTTACTG
 CCAATTTTTCTCCTCTTCATAACCATAAAAAGCTAGTATTGTAGAATCTTTATTGTTCCGAGCAGTGCGGCG
 CGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGT
 CGGAGGGCTGTGCGCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTA
 TTAAGTAAAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTACATTTCCACAACAT
 ATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTG
 CGTCCATCCAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGACTGCCGACAACAATAGTAT
 GCCCCATGGTGCAGTATCTAGTTACGCCAAATTAGTGCAAAACCAAACACCTGAAGACATTTTGGAAG
 AGTTTCTGAAATTATTCCATTACAACAAGACCTAATACCCGATCTAGTGAGACGTCAAATGACGAA
 AGCGGAGAAACATGTTTTTCTGGTCATGATGAGGAGCAAATTAAGTTAATGAATGAAAATTGTATTGT
 TTTGGATTGGGACGATAATGCTATTGGTGCCGGTACCAAGAAAGTTTGTCAATTAATGGAAAATATTG
 AAAAGGGTTTACTACATCGTGCAATTCTCCGTCTTTATTTTCAATGAACAAGGTGAATTACTTTTACAAC
 AAAGAGCCACTGAAAAAATAACTTTCCCTGATCTTTGGACTAACACATGCTGCTCTCATCCACTATGTA
 TTGATGACGAATTAGGTTTGAAGGGTAAGCTAGACGATAAGATTAAAGGGCGCTATTACTGCGGCGGTG
 AGAAAACCTAGATCATGAATTAGGTATTCCAGAAGATGAACTAAGACAAGGGGTAAGTTTCACTTTTT
 AAACAGAATCCATTACATGGCACCAAGCAATGAACCATGGGGTGAACATGAAATTGATTACATCCTAT
 TTTATAAGATCAACGCTAAAGAAAACCTTGACTGTCAACCCAAACGTCAATGAAGTTAGAGACTTCAAA
 TGGGTTTCACCAAATGATTTGAAAACCTATGTTTGCTGACCCAAGTTACAAGTTTACGCCTTGGTTTAAG
 ATTATTTGCGAGAATTACTTATTCAACTGGTGGGAGCAATTAGATGACCTTTCTGAAGTGGAATAATGA
 CAGGCAAATTCATAGAATGCTATAACAACGCGTCAATAATATAGGCTACATAAAAAATCATAATAACTT
 TGTTATCATAGCAAAATGTGATATAAAACGTTTTCATTTACCTGAAAAATAGTAAAAATAGGCGACAA

AAATCCTTAGTAAATATGTAAACCTTATTTCTTTATTTACCCGGGAGTCAGTCTGACTCTTGGGAGAGA
TGAGGATGTAATAATACTAATCTCGAAGATGCCATCTAATACATATAGACATACATATATATATATAT
ACATTCTATATATTCTTACCCAGATTCTTTGAGGTAAGACGGTTGGGTTTTATCTTTTGCAGTTGGTACT
ATTAAGAACAATCGAATCATAAGCATTGCTTACAAAGAATACACATACGAAATATTAACGATAATGTC
AATTACGAAGACTGAACTGGACGGTATATTGCCATTGGTGGCCAGAGGTAAAGTTAGAGACATATATG
AGGTAGACGCTGGTACGTTGCTGTTTGTGTGCTACGGATCGTATCTCTGCATATGACGTTATTATGGAAA
ACAGCATTCTGAAAAGGGGATCCTATTGACCAAACCTGTCAGAGTTCTGGTTCAAGTTCTGTCCAAC
GATGTTTCGTAATCATTTGGTCGACATCGCCCCAGGTAAGACTATTTTCGATTATCTACCTGCAAAATTG
AGCGAACCAAAGTACAAAACGCAACTAGAAGACCGCTCTCTATTGGTTCACAAACATAAACTAATTCC
ATTGGAAGTAATTGTGACAGGGCTACATCACCGGATCTGCTTGGAAAGAGTACGTAAAAACAGGTACTG
TGCATGGTTTGAAACAACCTCAAGGACTTAAAGAATCTCAAGAGTTCCCAGAACCAATCTTCACCCCA
TCGACCAAGGCTGAACAAGGTGAACATGACGAAAACATCTCTCTGCCAGGCCGCTGAGCTGGTGG
GTGAAGATTTGTCACGTAGAGTGGCAGAACTGGCTGTAAAACCTGTAAGTCCAAGTGCAAAGATTATGCT
AAGGAGAAGGGCATCATCATCGCAGACACTAAATTGTTTAAAC

SEQ ID NO: 53

pAM426 sequence

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTCCGGG
GCTGGCTTAACCTATGCGGCATCAGAGCAGATTGTAAGTGCAGAGTGCACCATATCGACTACGTCGTAAGG
CCGTTTCTGACAGAGTAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCTTCAAGAAGGTATT
GACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTTGTATTTTTTTTTTTTAGAGA
AAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTTTCTGTTACACCTAACTTTTGTGTGGTG
CCCTCCTCCTTGTCAATATTAATGTTAAAGTGCAATCTTTTTCTCTTATCACGTTGAGCCATTAGTATCA
ATTTGCTTACCTGTATTCCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATA
TAGTTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTTCGTGTCGTTTCTATTATGAATTTCAATTTAT
AAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGC
GACAGCATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCCA
AAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCA
TTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTTCTTTGGCAAATCTGGAGCAGAACCGTGG
CATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAAGGACGCAGATGGCAACA
AACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGAGATGATATCACCAAACATGTTGCTGGTGATT
ATAATACCATTAGGTGGGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTG
AACCTTCAATGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGC
CAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGG
CCATTCTTGTGATTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCAT
CGTCTTCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTT
TAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAG
TTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAAGTCTAACACTACCGGTACCC
CATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCTTGGAGGCTTCCAGCGCCTCATC
TGGAAGTGGGACACCTGTAGCATCGATAGCAGCACCAACCAATTAAATGATTTTCGAAATCGAACTTGA
CATTGGAACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACG

TGGTCACCTGGC⁷⁰⁰⁰AAAACGACGATCTTCTTAGGGGCAGACATTACAATGGTATATCCTTGAAATATATA
 TAAAAAAGGCGCCTTAGACCGCTCGGCCAAACAACCAATTACTTGTTGAGAAATAGAGTATAATTAT
 CCTATAAATATAACGTTTTTTGAACACACATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAAT
 GTGGATTTTGATGTAATTGTTGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACT
 AGAAGTTCTCCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGC
 ATCAGGAAATTGTAAACGTTAATATTTTGTAAAAATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTT
 TAACCAATAGGCCGAAATCGGCCAAAATCCCTTATAAATCAAAAAGAAATAGACCGAGATAGGGTTGAGT
 GTTGTTCAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAAC
 CGTCTATCAGGGCGATGGCCCACTACGTGGAAGATCCGAGGCCTAGCTTTAACGAACGCAGAATTTTC
 GAGTTATTAACTTAAAAATACGCTGAACCCGAACATAGAAATATCGAATGGGAAAAAAAAAACTGCAT
 AAAGGCATTAAAAAGAGGAGCGAATTTTTTTTTTAATAAAAAATCTTAATAATCATTAAAAAGATAAATAAT
 AGTCTATATATACGTATATAAATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTAGCGTAAAG
 GATGGGGAAAGAGAAAAAGAAAAAAATTGATCTATCGATTTCAATTCAATTCAATTTATTTCTTTTCGG
 ATAAGAAAGCAACACCTGGCAATTCCTTACCTTCCAATAATTCCAAAGAAGCACCACCACCAGTAGAG
 ACATGGGAGACCCGGGCCATGGTTAGATAGACATAGGGTAACTAGCAATGATTTGATCAAATGCTTG
 TATTCATCTCCCATTTCTCGTAAAATTGTCTTTACCTGCATATTGGACCTCTAAAAATTGGCAAAGATAT
 ATAACAGCCATAAGTAAAGGTCTTGGGATATTCTTTGTTGTTAAATACTCTCTGTTTATGTCTTTCCAA
 ACGTCCTCCACTTCCTTATAAATCAGTGTCTGAGCATATTCTTCGTTGACATTGTATTCCTTCATGTAAG
 ATTCTAAAGAGCTTGAACCTATGTTTTCTCTCCTGTTCCGCTTTATGAGTCATCAGGTCATTTAATCTCCT
 ACCCAGAATACCACTGTAACGGAATAAAGGCGGAGCAGATACAGCCCACTCAACTGATTCCTTAGTGA
 AAATATCGCTCATTCTAGATAACAGGTAGTTGTTAGCAAGTTTGCACCACCAGTGATAATAACTACG
 GGATCGTGCTCTTCAGTTGTGCGGTATGTGTCCTTCATTAGCCCATTTTCGCTTCTACCATTAGATTCCTTA
 CGAATTCTTTAACGAACCTCTTCCACAGTTGAATAAATCAGTTCTACCTTCTTTGGCCAGAACTCCT
 CCATTTCTGTGTAGGTATCCATGAATAATTTGTAAATAGGCTTCATGTATTCCGGCAACGTGTCTAAGC
 AGGTGATCGACCATCTTTCCACGGCTTCAGTGAAAATCTTTAACTCCTCGTAAGTTCCATATGCGTCAT
 ACGTGTCAATCAATAAGTGTTATCACAGCAACTGCCTTAGTGAAAAAACTCTAGCTCTTGAATACTGG
 GGTTCTGTAACCAGAACCTAAACCCCAAAAATAGCATTCAACGATACGATCTCTCAGACATGGGGCATT
 TTTCTTAATATCAAATGCCTTCCACCACCTTGCATACGTGACTCAACTCTTCTTATGTAGGCTCTGCAAT
 AGATTGAACTCCAGTTTAGCTAACTTTAGCAGAGTTTTATTATGGGAGTCTTGTTGCTGATAGAAGGGT
 ATGTACTGGGCGGCCTCGATCCTTGGCAATCTCTTCCACAATGGTTGCTTTAAAGCTCTCTGGATTTCA
 GTGAATAAAGCGGGGTTTGTACTAAACGCGTCCTTTGTCATAATCGATAGCCTTGATCTTGTGAATCCC
 AGGGCATCTTCAAGAATTATTTGCCCCGGAACCTCTCATGGACGTAGCCTCATATAATTCCAACAATCCT
 TCAACATCATTGCTAACGATTGTTTAAAAGCACCATTCTTGTCTTTATAGTTATTAAACACATCACAC
 GTGACATAGTATCCTTGTTTACGCATCAGCCTAAACCATAAGCTAGACCTGTGCGCCATTCCAATTATCA
 CCATAGGTCTCGTAAATACATTGCAATGCATGATCAATTTACAGTTCAAATGATACGGAATACCTAA
 ACGTTGAATCTCGTCAATCAGCTTCAACAAATTTGCATGTTTCATAGGAATATCCAATGCTTCCTTTAA
 CAACTGTCTTACTTCCTTCTTTAGATCGTTTACTATTTGCTCCACACCCCTGTTCAACTTGTTTCTCATAA
 ATCAAAAATTGATCGCCCCAAATAGAAGGTGGGAAATTTGCAATTGGCCTTATAGGTTTCTCTTCAGTC
 AAGGCCATTGTTTTCTGCAGATCCGGGGTTTTTTCTCCTTGACGTAAAGTATAGAGGTATATTAACAA
 TTTTTGTTGATACITTTTATTACATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGT
 TAAAGTGGTTATGCAGTTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTA
 ATTACCCAGAAATAAGGCTAAAAAACTAATCGCATTATCATCCTATGGTTGTTAATTTGATTCGTTCA

TTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCCTTCATAACCATAAAAAGCTAGTATTGTAG
 AATCTTTATTGTTTCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGAC
 GAGGACGCACGGAGGAGAGTCTTCCTTCGGAGGGGCTGTACCCCGCTCGGCGGGCTTCTAATCCGTACTA
 AGATCTGCTTTAATTTGGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAAGGAGCGGGC
 GCTAGGGCGCTGGCAAGTGTAGCGGTACAGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCC
 GCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG
 CCTCTTCGCTATTACGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGCTATT
 GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCA
 GCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG
 CAAAAGGCCAGCAAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
 CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
 AGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT
 CGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCC
 TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCAC
 TGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT
 ACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA
 GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAG
 ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG
 GAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTT
 TAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAA
 TGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCC
 GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
 CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAAGT
 GGTCTTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG
 CCAGTTAATAGTTTTCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGT
 ATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAA
 AGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGT
 TATGGCAGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTA
 CTCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGG
 ATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
 CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCACTGATCTTCA
 GCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
 AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCA
 GGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGC
 GCACATTTCCCCGAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTAGAACAAAAATGCAA
 CGCGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGA
 AAGCGCTATTTTACCAACGAAGAATCTGTGCTTCATTTTTGTAAACAAAAATGCAACGCGAGAGCGC
 TAATTTTTCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTT
 TACCAACAAAGAATCTATACTTCTTTTTTGTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCTAAC
 AAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAACTTTTTGCAC
 TGTAGGTCCGTAAAGGTTAGAAGAAGGCTACTTTGGTGTCTATTTTTCTCTCCATAAAAAAAGCCTGAC

TCCACTTCCC¹CG²GT³TTACTGATTACTAGCGAAGCTGCGGGTGCATTTTTTCAAGATAAAAGGCATCCCCG
ATTATATTCTATACCGATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTTC
ATTGGTCAGAAAATTATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAATGTTTACA
TTTTCGTATTGTTTTCGATTCACTCTATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACT
AGAGATAAACATAAAAAATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGG
TAGGTTATATAGGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAG
CGGTATTCGCAATATTTTAGTAGCTCGTTACAGTCCGGTGC⁴GT⁵TTTTTGGTTTTTTGAAAGTGC⁶GT⁷CTTCA
GAGCGCTTTTGGTTTTCAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAG
GAACTTCAAAGCGTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCGAGCTGCGCACATACAGCTCA
CTGTTACGTCGCGACCTATATCTGCGTGTGCTGTATATATATATACATGAGAAGAACGGCATAGTGC
GTGTTTATGCTTAAATGCGTACTTATATGCGTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCT
GTGATATTATCCCATTCATGCGGGGTATCGTATGCTTCCTTCAGCACTACCTTTTAGCTGTTCTATATG
CTGCCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCCTTTGATATTGGATCATACTAA
GAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCTTTCGTC

SEQ ID NO: 54

pAM322 sequence

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGG
GCTGGCTTAACTATGCGGCATCAGAGCAGATTGTA⁸CTGAGAGTGCACCATATCGACTACGTCGTAAGG
CCGTTTCTGACAGAGTAAAATTCTTGAGGGA⁹ACTTTCA¹⁰CCATTATGGGAAATGCTTCAAGAAGGTATT
GACTTAAACTCCATCAAATGGTCAGGTCATTGAGTGT¹¹TTTTATTTGTTGTATTTTTTTTTTTTAGAGA
AAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTTTCTGTTACACCTAACTTTTTGTGTGGT
CCCTCCTCCTTGTCAATATTAATGTTAAAGTGCAATTCTTTTCCCTTATCACGTTGAGCCATTAGTATCA
ATTTGCTTACCTGTATTCCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATA
TAGTTTCGTCTACCCTATGAACATATTCATTTTGTAATTCGTGTCGTTTCTATTATGAATTCATTTAT
AAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTAAAGCAAGGATTTCTTAACTTCTTCGGC
GACAGCATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATACCTGCATCCA
AAACCTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCATGACAATTTCAACATCA
TTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGCAGAACCGTGG
CATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAAGGACGCAGATGGCAACA
AACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGAGATGATATCACCAAACATGTTGCTGGTGATT
ATAATACCATTTAGGTGGGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTG
AACCTTCAATGTAGGGAATTCGTTCTTGATGGTTTCTCCACAGTTTTTCTCCATAATCTTGAAGAGGC
CAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGG
CCATTCTTGTGATTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCAT
CGTCTTCCTTTCTCTTACCAAAGTAAATACCTCCCCTAATTCTCTGACAACAACGAAGTCAGTACCTT
TAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAAG
TTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAAGTCTAACACTACCGGTACCC
CATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCTTGGAGGCTTCCAGCGCCTCATC
TGGAAGTGGGACACCTGTAGCATCGATAGCAGCACCACCAATTAATGATTTTCGAAATCGAACTTGA
CATTGGAACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTGACCAACG

TGGTCACCTGGC⁻⁻⁻AAAACGACGATCTTCTTAGGGGCAGACATTACAATGGTAA⁻⁻⁻ATCCTTGA⁻⁻⁻AA⁻⁻⁻ATATA
 TAAAAAAGGGCGCCTTAGACCGCTCGGCCAAACAACCAATTACTTGTTGAGAAATAGAGTATAATTAT
 CCTATAAATATAACGTTTTTTGAACACACATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAAT
 GTGGATTTTGATGTAATTGTTGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGGATATACT
 AGAAGTTCTCCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGC
 ATCAGGAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTT
 TAACCAATAGGCCGAAATCGGCCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGT
 GTTGTTCAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAAC
 CGTCTATCAGGGCGATGGCCCACTACGTGGAAGATCCGAGGGCCTAGCTTTAACGAACGCAGAATTTTC
 GAGTTATTA⁻⁻⁻AACTTAA⁻⁻⁻AATACGCTGA⁻⁻⁻ACCCGAACATAGAAATATCGAATGGGAAAA⁻⁻⁻AAAACTGCAT
 AAAGGCATTAA⁻⁻⁻AGAGGAGCGAATTTTTTTTTTAATAAAAAATCTTAATAATCATTA⁻⁻⁻AGATAAATAAT
 AGTCTATATATACGTATATAAATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTAGCGTAAAG
 GATGGGGAAAGAGAAAAAGAAAAAAATTGATCTATCGATTTCAATTCAATTCAATTTATTTCTTTTCGG
 ATAAGAAAGCAACACCTGGCAATTCCTTACCTTCCAATAATTCCAAGAAGCACCACCAGTAGAG
 ACATGGGAGACCCGGGCCATGGTTAGATAGACATAGGGTAAACTAGCAATGATTTGATCAAATGCTTG
 TATTCATCTCCATTCTCGTAA⁻⁻⁻AATTGTCTTTACCTGCATATTGGACCTCTAA⁻⁻⁻AAATTGGCAAAGATAT
 ATAACAGCCATAAGTAAAGGTCTTGGGATATTCTTTGTTGTTAAATACTCTCTGTTTATGTCTTTCCAA
 ACGTCCCTCCACTTCCTTATAAATCAGTGTCTGAGCATATTCTTCGTTGACATTGTATTCCTTCATGTAAG
 ATTCTAAAGAGCTTGA⁻⁻⁻ACTATGTTTTCTCTCCTGTTCCGCTTTATGAGTCATCAGGTCATTTAATCTCCT
 ACCCAGAATACCACTGTAACGGAATAAAGGCGGAGCAGATACAGCCCACTCAACTGATTCCTTAGTGA
 AAATATCGCTCATTCTAGATAACAGGTAGTTGTTAGCAAGTTTGCACCACCAGTGATAATAACTACG
 GGATCGTGCTCTTCAGTTGTCTGGTATGTGTCTTCATTAGCCCATTTTCGCTTCTACCATTAGATTCCTTA
 CGAATTCTTTAACGAACTCCTTCCCACAGTTGAATAAATCAGTTCTACCTTCTTTGGCCAGAACTCCT
 CCATTTCTGTGTAGGTATCCATGAATAATTTGTAAATAGGCTTCATGTATTCCGGCAACGTGTCTAAGC
 AGGTGATCGACCATCTTTCCACGGCTTCAGTGAAAATCTTTAACTCCTCGTAAGTTCCATATGCGTCAT
 ACGTGTCAATCAATAAGTGTTATCACAGCAACTGCCTTAGTGAAAAAACTCTAGCTCTTGAATACTGG
 GGTTCTGTAACCAGAACCTAAACCCCCAAAAATAGCATTCAACGATACGATCTCTCAGACATGGGGCATT
 TTTCTTAATATCAAATGCCTTCCACCCTTGCATACGTGACTCAACTCTTCCTTATGTAGGCTCTGCAAT
 AGATTGA⁻⁻⁻ACTCCAGTTTAGCTAACTTTAGCAGAGTTTTATTATGGGAGTCTTGTGCTGATAGAAGGGT
 ATGTACTGGGCGGCCTCGATCCTTGGCAATCTCTTCCACAATGGTTGCTTTAAAGCTCTCTGGATTTCA
 GTGAATAAAGCGGGGTTTGTACTAAACGCGTCCTTTGTCATAATCGATAGCCTTGATCTTGTGAATCCC
 AGGGCATCTTCAAGAATTATTTGCCCCGGA⁻⁻⁻ACTCTCATGGACGTAGCCTCATATAATTCCAACAATCCT
 TCAACATCATTCGCTAACGATTGTTTTAAAGCACCATTCTTGTCTTTATAGTTATTAACACATCACAC
 GTGACATAGTATCCTTGTTACGCATCAGCCTAAACCATAAGCTAGACCTGTCGCCATTCCAATTATCA
 CCATAGGTCTCGTAAATACATTGCAATGCATGATCAATTTACGTTCAA⁻⁻⁻AATGATACGGAATACCTAA
 ACGTTGAATCTCGTCAATCAGCTTCAACAAATTTGCATGTTTCATAGGAATATCCAATGCCTTCCTTAA
 CAACTGTCTTACTTCCTTCTTTAGATCGTTTACTATTIGCTCCACACCCTGTTCAACTTGTTCCTCATAA
 ATCAAAAATTGATCGCCCCAAATAGAAAGGTGGGAAATTTGCAATTGGCCTTATAGGTTTCTCTTCAGTC
 AAGGCCATTGTTTTCTGCAGATCCGGGGTTTTTTCTCCTTGACGTAAAGTATAGAGGTATATTAACAA
 TTTTTGTTGATACTTTTATTACATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGT
 TAAAGTGGTTATGCAGTTTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTA
 ATTACCCAGAAATAAGGCTAAAAAACTAATCGCATTATCATCCTATGGTTGTTAATTTGATTCGTTCA

TTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCTCTTCATAACCATAAAAAGCTAGTATTGTAG
AATCTTTATTGTTTCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGAC
GAGGACGCACGGAGGAGAGTCTTCCTTCGGAGGGGCTGTCACCCGCTCGGCGGCTTCTAATCCGTACTA
AGATCTGCTTTAATTTGGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGC
GCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCGCGCTTAATGCGCC
GCTACAGGGCGCGTCGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGG
CCTCTTCGCTATTACGCCAGCTGAATTGGAGCGACCTCATGCTATACCTGAGAAAGCAACCTGACCTA
CAGGAAAGAGTACTCAAGAATAAGAATTTTTCGTTTTAAAACCTAAGAGTCACTTTAAAATTTGTATA
CACTTATTTTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCTTATTTAGAAG
TGTCACAACGTATCTACCAACGATTTGACCCTTTTCCATCTTTTCGTAAATTTCTGGCAAGGTAGACA
AGCCGACAACCTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAATTGTTAATTAAGAGTCAGTCGA
CTTAAAACTAGGGACCAATAGCAATTCTGTTTTACGTTGCATTGTTGCACCTGAACCTTCCGTCATGT
CAATTTGATCATATGAAACTCCATTGGGCAACTTCCAGTTGAAATGATAAAGAATGTTGGCTAGTGGC
AGTTGAACATTGGCCAAACCTAACGCAGCGCCAGGACACATACGACGTCCAGCCCCAAATGGTAAAT
ATTCATATTCGGCGCCCATCACTGTTGCCGAAGAGTTTTCAAATCTTTCAGGTATAAACGCTTCTGCAT
CCTTCCAGTATTCAGGATCTCTATTGATCGCAAACACATTAACGATTAATTTGTTTTGTTAGGGATAT
TATAACCAGCCAAGTTTACTGGCTGACGACATTCTCTAGGTAGCACTAACGGCAAGGGTGGGTGTAGT
CTAAGAGTCTCTTTGATGACCATATTCAAGTAGGACAATTCTTGTATATCTTCTTCATGTATTTTTCTT
TCCCATTCAAGGCCTTACGTAATTCAGCCTGAACCTTTTTCCATTGCTTTCGGACATTTTATTAGCTCGCT
TATAGCCCATTTCTATGGTAGAACTTGAAGTGTGGTCCCTGCACCGAACATGTCCAAAATTATTGCTTT
GATATTATCCGAAGTCAGAGGAACTCAGCAGAATCCTTTAATCTAAGTAATACATCTAATAGGGTTT
CGTTGGTTTTGGATGACGTATTTACGGTATGTTTCAGCTACCAAATTGTCAATTAAGTTATCAATCTTTTT
ACGTAGGCTAGTTAATCTTGCTCTCTTACCGCTCAAGTGATGCAAGAACTTTTTAGATGGGAAAATATC
GGCAACATCGAAACCGCCTGTTTGTCTCAGTATTTCTTTAACAATTTAGTAAGTTCCTTTTGATCTTTA
ATTCCCTTACCAAACGCAGCACGGGATAGTATAGTGGCAATTAGTTTAAAAACGTTTTCACTTAAATTT
ACTGGTCTACCACTACCTGAAGCCTTTATTTCTGGACTAAATTCCAACATTCTTCTTCCCTCAACGATT
GAAATGACTTAACCTTTTTTACAGACAACAATCAAGAGTACAAATCTTCCCTTAATTGTCTCCAGTATT
CCCCATATGGAGCAAGGACAACATCAGTGTTATGATATAAACTATTTCCCCAGTTAAAGTTTCGGGT
CTATTAGCGAAAGTAATATCGTAGGTTGTAAGAAFTTCCTTAGCCCACTTAGGACTCGACACGACTATT
GTGGGTACCTCTCCCAATTGAAGGTGCATTAGCGAACCATATTTTCTCGCTAAATCCCTTACACCCCTG
TGTGGTGTGGTTCCGATCAAATGGTGCATGTGACCAATGATGGGTAGCCTCCAAGGTTCGGCAAGGA
CTTTTAGTTGACTTACTTCTAGTGGCAAATTTGTACACGAACAACAAAATAGTTGCTAAAGCAATTGA
TGTAAGTTAAAGATAGTGCCATAGCCTTTAAAAATTGACTTCATTGTTTTCTAGGCCTTTAGTGAGGGTT
GAATTCGAATTTTCAAAAATTCTTACTTTTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTAC
ATGGCATTACCACCATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAAT
GTAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACCTTCAGTAATACGCTTAACTGC
TCATTGCTATATTGAAGTACGGATTAGAAGCCGCGGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTC
CTCCGTGCGTCCTCGTCTTCACCGGTGCGGTTCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCG
AACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCC
CACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCT
TATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAA
CTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAA

AGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACTGCAAGGAGAAAAAACCCTCAAGCTTCC
CGGGAAAAACAATGCAATCGACAACCTTCCGTTAACTATCACCTTTCGATCTTATGACTGCCTTGTAAAA
TGGTAAAGTTAGTTTCGACACGTCCAATACTTCCGATACAAATATACCACTGGCGGTTTTTCATGGAAA
ACAGGGAATTGCTTATGATATTAACAACCAGTGTGGCCGTTTTTAATTGGTTGTGTGGTTGTATTGGTAT
GGAGAAGATCATCAAGTGCCGCTAAGAAGGCCGCCGAATCACCAGTCATTGTCTGCCAAAGAAAGT
CACTGAAGATGAGGTTGATGACGGCAGAAAGAAAGTTACTGTATTTTTTCGGGACACAAACGGGGACT
GCGGAAGGTTTTGCGAAAGCTCTAGTTGAAGAAGCCAAGGCAAGGTACGAAAAAGCAGTATTCAAAG
TTATTGATTTAGATGACTACGCCGAGAAAGATGATGAATACGAAGAAAAGCTAAAGAAAGAATCTTTG
GCATTCTTCTTTTTAGCTACCTATGGTGACGGAGAACCAACAGATAACGCCGCTAGATTCTATAAATGG
TTTACTGAAGGAGAAGAAAAAGGTGAGTGGTTAGATAAGTTACAATACGCTGTCTTTGGATTGGGAAA
TCGTCAATATGAACACTTCAATAAGATTGCAAAAGTGGTCGATGAAAAATTAGTTGAGCAGGGGGCTA
AAAGGTTAGTGCCTGTGCGTATGGGTGATGACGATCAATGTATCGAAGATGATTTTACTGCTTGGAAG
GAATTGGTTTGGCCAGAATTAGATCAGCTATTGAGGGACGAAGATGACACAAGTGTCTGCTACTCCGTA
CACCGCCGCTGTTGGCGAATATCGTGTTGTTTTTCACGATAAACCTGAACTTACGATCAAGATCAATT
GACCAACGGACACGCAGTTCACGACGCCCAACACCCATGCAGATCGAACGTTGCGGTCAAGAAAGAA
TTACACAGTCCCTTATCCGATAGGAGTTGTACTCATTTAGAATTTGATATTTCCAATACTGGACTATCG
TATGAAACTGGCGACCATGTCTGGTGTATATGTGGAAAACCTGTCTGAAGTTGTAGATGAAGCCGAAAA
ATTGATTGGGCTTCCTCCACATACATACTTTTCTGTGCATACAGATAATGAAGATGGTACTCCACTTGG
CGGAGCCTCGTTACCACCTCCCTTTCCACCATGTACACTTAGAAAAGCTCTTGCATCTTATGCAGATGT
ACTTTCTTACCAAAGAAAAAGTGCATTACTAGCTCTAGCCGCCCATGCTACCGACTCTACTGAAGCTGA
CCGTTTGAAATTCTTTGCTTCACCTGCTGGCAAAGACGAGTACGCACAGTGGATTGTGGCATCTCACAG
ATCATTGCTGGAAGTGATGGAAGCCTTCCCATCGGCAAAGGCCACCATTAGGCGTGTTTTTCGCATCTGT
TGCCCCACGTTTACAGCCTAGATACTATTCCATATCTTCTAGCCCCAAAATTTGCCCCCAATCGTATTCA
TGTGACGTGTGCGCTGGTGTATGAACAACTCCATCAGGAAGGGTACATAAAGGTGTCTGTAGTACAT
GGATGAAAAACGCGGTGCCAATGACTGAATCTCAAGATTGTTTCGTGGGCACCAATTTATGTTTCGTACT
TCTAATTTTAGACTACCTAGTGACCCTAAAGTACCAGTGATTATGATCGGGCCTGGGACAGGACTAGC
GCCATTACAGAGGTTTCTTACAAGAAAGATTGGCCCCAAAAGGAAGCAGGTACGGAATTAGGAACCGCA
ATTCTATTCTTTGGTTGTCGTAATAGAAAAGTTGACTTTATATACGAAGATGAGTTAAACAACCTTCGTT
GAACTGGAGCGTTATCAGAATTAGTGACAGCATTTCTTAGGGAAGGTGCAACAAAAGAATACGTCC
AACATAAAATGACCCAAAAGGCCAGCGATATATGGAATTTGCTGTCCGAGGGTGCCTATTTGTACGTT
TGTGGTGATGCAAAGGGAATGGCTAAAGATGTTACAGGACATTGCATACAATTGTTACAGGAACAAG
GTTCCCTGGATTCTCTAAGGCAGAACTTTATGTTAAAAACCTTCAGATGGCTGGTAGATATTTGCGTG
ATGTTTGGTGAGCTAGCTAAGATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGG
TCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTCTTTTTTTT
CTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTC
GAAGATCCAGCTGCATTAATGAATCGGCCAACCGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCTGCTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
AGGCGGTAATACGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
GTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGT

CGTTCGCTCCAAAGCTGGGCTGTGTGCACGAACCCCCGTTACAGCCGACCGCTGCGCTTATCCGGTAA
 CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGA
 TTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT
 AGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTC
 TTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
 GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAAAC
 TCACGTAAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
 TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
 GAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
 ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC
 GGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
 TTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCCGCAGTTAATAGT
 TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTC
 AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTC
 CTTCCGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACT
 GCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
 ATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGC
 CACATAGCAGAACTTTAAAAGTGCTCATCTTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC
 TTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTTACT
 TTCACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
 CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
 CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
 GAAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTAGAACAAAAATGCAACGCGAGAGCGCT
 AATTTTTCAAACAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTT
 ACCAACGAAGAATCTGTGCTTCATTTTTGTAAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAA
 CAAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATTTTACCAACAAAGA
 ATCTATACTTCTTTTTTGTCTACAAAAATGCATCCCGAGAGCGCTATTTTTCTAACAAAGCATCTTAG
 ATTACTTTTTTCTCCTTTGTGCGCTCTATAATGCAGTCTCTTGATAACTTTTTGCACTGTAGGTCCGTTA
 AGGTTAGAAGAAGGCTACTTTGGTGTCTATTTTCTCTCCATAAAAAAGCCTGACTCCACTTCCCGCG
 TTTACTGATTACTAGCGAAGCTGCGGGTGCATTTTTTCAAGATAAAGGCATCCCCGATTATATTCTATA
 CCGATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAA
 ATTATGAACGGTTTCTTCTATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTT
 TTCGATTCACTCTATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACAT
 AAAAAATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGTAGGTTATATAG
 GGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCAATGTTTGTGGAAGCGGTATTTCGAA
 TATTTTAGTAGCTCGTTACAGTCCGGTGCGTTTTTGGTTTTTGAAGTGCGTCTTCAGAGCGCTTTTGG
 TTTTCAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTCAAAG
 CGTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCGAGCTGCGCACATACAGCTCACTGTTACGTC
 GCACCTATATCTGCGTGTGCTGTATATATATATACATGAGAAGAACGGCATAGTGCGTGTATTATGCT
 TAAATGCGTACTTATATGCGTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATC
 CCATTCCATGCGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCT

CAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCTTTGATATTGGATCATACTAAGAAACCATTAT
TATCATGACATTAACTATAAAAAATAGGCGTATCACGAGGCCCTTTCGTC

Claims

1. A method for producing an isoprenoid comprising:

culturing a microorganism in a medium with a limited carbon source that provides for 75% or less of a maximum specific growth rate of the microorganism, wherein the maximum specific growth rate is a rate that would have been achieved by culturing the microorganism at an optimal temperature for growth in a medium in which nutrients are present in excess,

wherein the microorganism comprises a heterologous nucleic acid encoding one or more enzymes of a deoxyxylulose 5-phosphate (DXP) pathway for making isopentenyl pyrophosphate, under control of at least one heterologous transcriptional regulator, and

wherein the heterologous nucleic acid encodes at least one enzyme selected from the group consisting of:

(i) an enzyme that condenses pyruvate with D-glyceraldehyde 3-phosphate to form 1-deoxy-D-xylulose-5-phosphate;

(ii) an enzyme that converts 1-deoxy-D-xylulose-5-phosphate to 2C-methyl-D-erythritol-4-phosphate;

(iii) an enzyme that converts 2C-methyl-D-erythritol-4-phosphate to 4-diphosphocytidyl-2C-methyl-D-erythritol;

(iv) an enzyme that converts 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate;

(v) an enzyme that converts 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate;

(vi) an enzyme that converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate; and,

(vii) an enzyme that converts 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate to one of IPP or DMAPP.

2. The method of claim 1 wherein the at least one heterologous transcriptional regulator is inducible.

3. The method of claim 1 wherein the DXP pathway enzyme is under control of a single transcriptional regulator.
4. The method of claim 1 wherein the DXP pathway enzyme is under control of multiple transcriptional regulators.
5. The method of claim 1 wherein the microorganism comprises a plurality of heterologous nucleic acids encoding all of the enzymes of a DXP pathway.
6. The method of claim 1 wherein the microorganism is *E. coli*.
7. The method of claim 1 wherein the microorganism is *S. cerevisiae*.
8. The method of claim 1 wherein the heterologous nucleic acid comprises a nucleic acid sequence encoding a DXP pathway enzyme from a bacterium having an endogenous DXP pathway.
9. The method of claim 8 wherein the bacterium is of a genus selected from *Enterococcus*, *Pseudomonas*, and *Staphylococcus*.
10. The method of claim 1 wherein the heterologous nucleic acid comprises a nucleic acid sequence encoding a DXP pathway enzyme selected from 1-deoxy-D-xylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate reductoisomerase, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase, and 4-diphosphocytidyl-2C-methyl-D-erythritol synthase.
11. The method of claim 1 wherein the heterologous nucleic acid comprises a nucleic acid sequence encoding a DXP pathway enzyme selected from 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase, and isopentenyl/dimethylallyl diphosphate synthase.
12. The method of claim 1 wherein the microorganism is cultured in a growth medium having both nutrients and temperatures maintained at levels below those that provide for the maximum specific growth rate of the host cells.

13. The method of claim 1 wherein the temperature of the medium is at least 2-20° C below the optimal temperature.
14. The method of claim 13 wherein the microorganism is cultured at a temperature at least 5° C below the optimal temperature.
15. The method of claim 13 wherein the temperature of the medium is at least 10° C below the optimal temperature.
16. The method of claim 1 wherein the medium provides for about 60% or less of the maximum specific growth rate.
17. The method of claim 1 wherein the medium provides for about 50% or less of the maximum specific growth rate.
18. The method of claim 1 wherein the medium provides for about 40% or less of the maximum specific growth rate.
19. The method of claim 1 wherein the medium provides for about 25% or less of the maximum specific growth rate.
20. The method of claim 1 wherein the medium provides for about 75%-10% of the maximum specific growth rate.
21. The method of claim 1 wherein the medium is nitrogen-restricted.
22. The method of claim 1 wherein the isoprenoid is produced in an amount greater than about 10 grams per liter of medium.
23. The method of claim 1 wherein the isoprenoid is produced in an amount greater than about 50 mg per gram of dry cell weight.
24. The method of claims 22 or 23 wherein the amount of isoprenoid is produced in less than about 150 hours.

25. The method of claims 22 or 23 wherein the amount of isoprenoid is produced in less than about 96 hours.
26. The method of claims 22 or 23 wherein the amount of isoprenoid is produced in less than about 72 hours.
27. The method of claim 1 wherein the isoprenoid is selected from the group consisting of a hemiterpene, monoterpene, diterpene, triterpene, tetraterpene, sesquiterpene, and polyterpene.
28. The method of claim 1 wherein the isoprenoid is a sesquiterpene.
29. The method of claim 1 wherein the isoprenoid is a C₅-C₂₀ isoprenoid.
30. The method of claim 1, wherein the nutrients comprise a carbon source and a nitrogen source.
31. The method of claim 1, wherein said one or more enzymes is a bacterial enzyme.
32. The method of claim 1, wherein said one or more enzymes is an *E. coli* enzyme.
33. The method of claim 1 wherein the isoprenoid is selected from the group consisting of abietadiene, amorphadiene, carene, α -farnesene, β -farnesene, farnesol, geraniol, geranylgeraniol, isoprene, linalool, limonene, myrcene, nerolidol, ocimene, patchoulol, β -pinene, sabinene, γ -terpinene, terpinene and valencene.
34. The method of claim 33 wherein the sesquiterpene is α -farnesene.
35. The method of claim 33 wherein the sesquiterpene is β -farnesene.
36. The method of claim 33 wherein the sesquiterpene is amorphadiene.
37. The method of claim 33 wherein the sesquiterpene is farnesol.
38. The method of claim 33 wherein the sesquiterpene is nerolidol.
39. The method of claim 33 wherein the sesquiterpene is patchoulol.

40. The method of claim 33 wherein the sesquiterpene is valencene.

Figure 1A

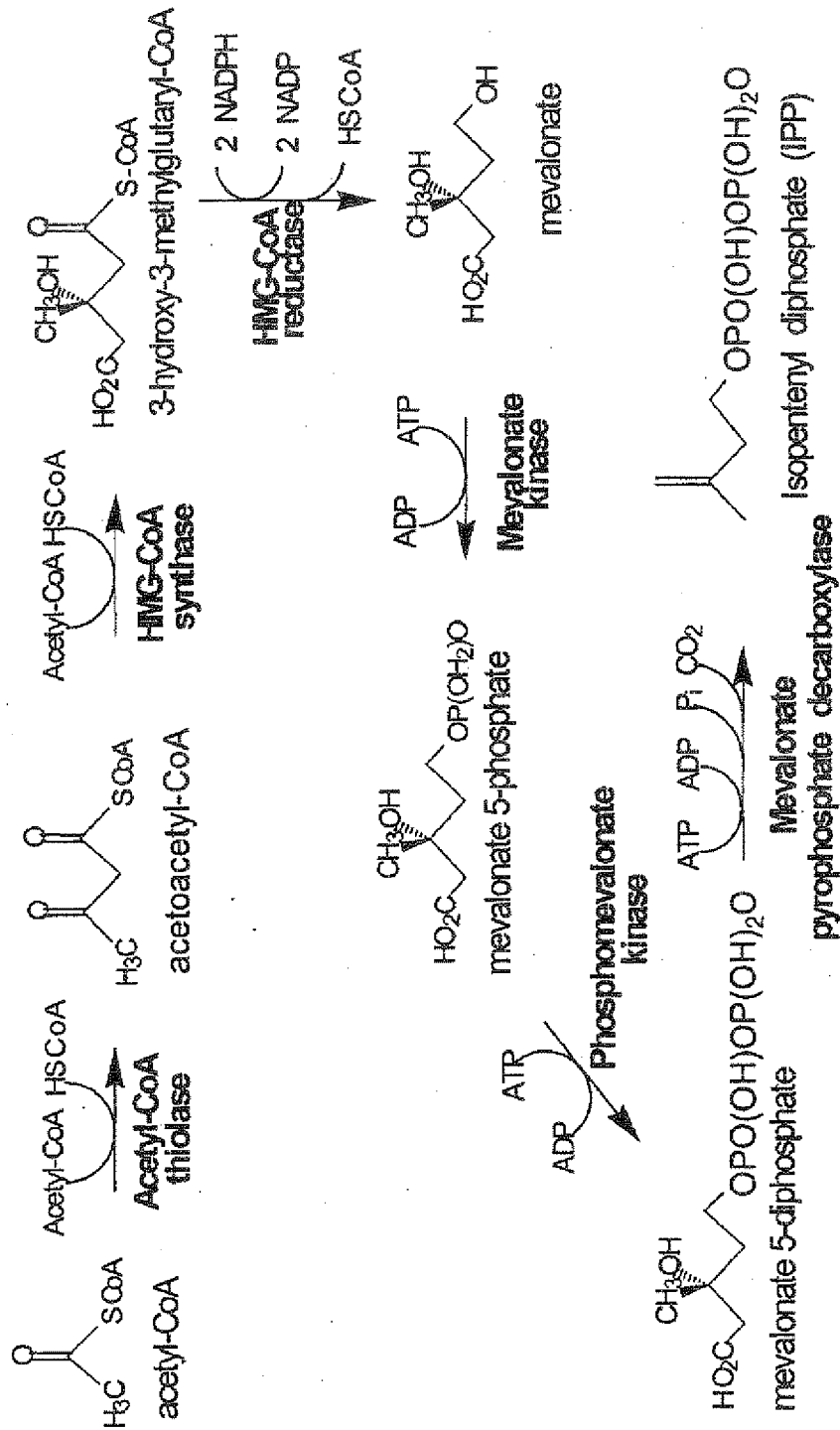


Figure 1B

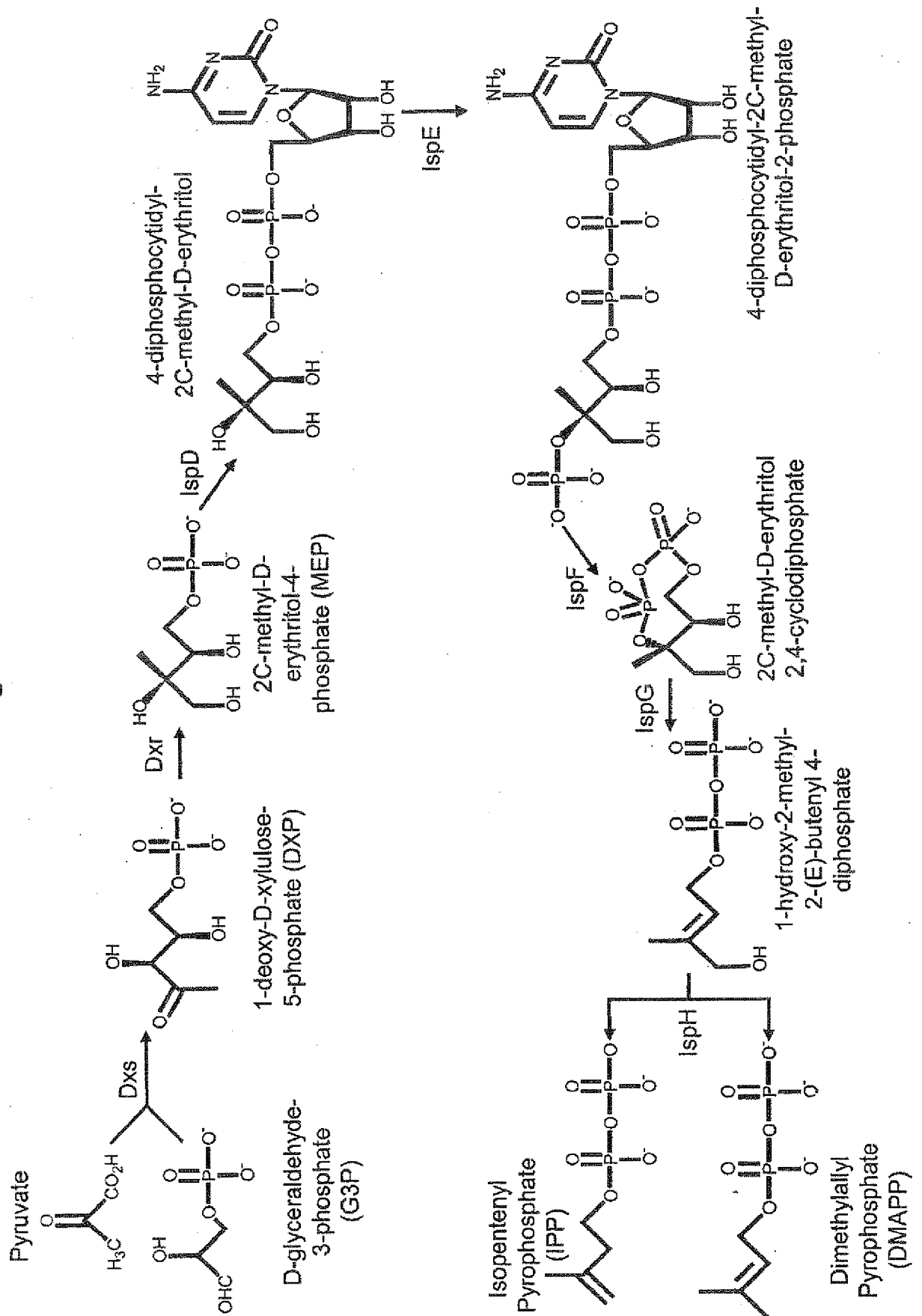


Figure 2

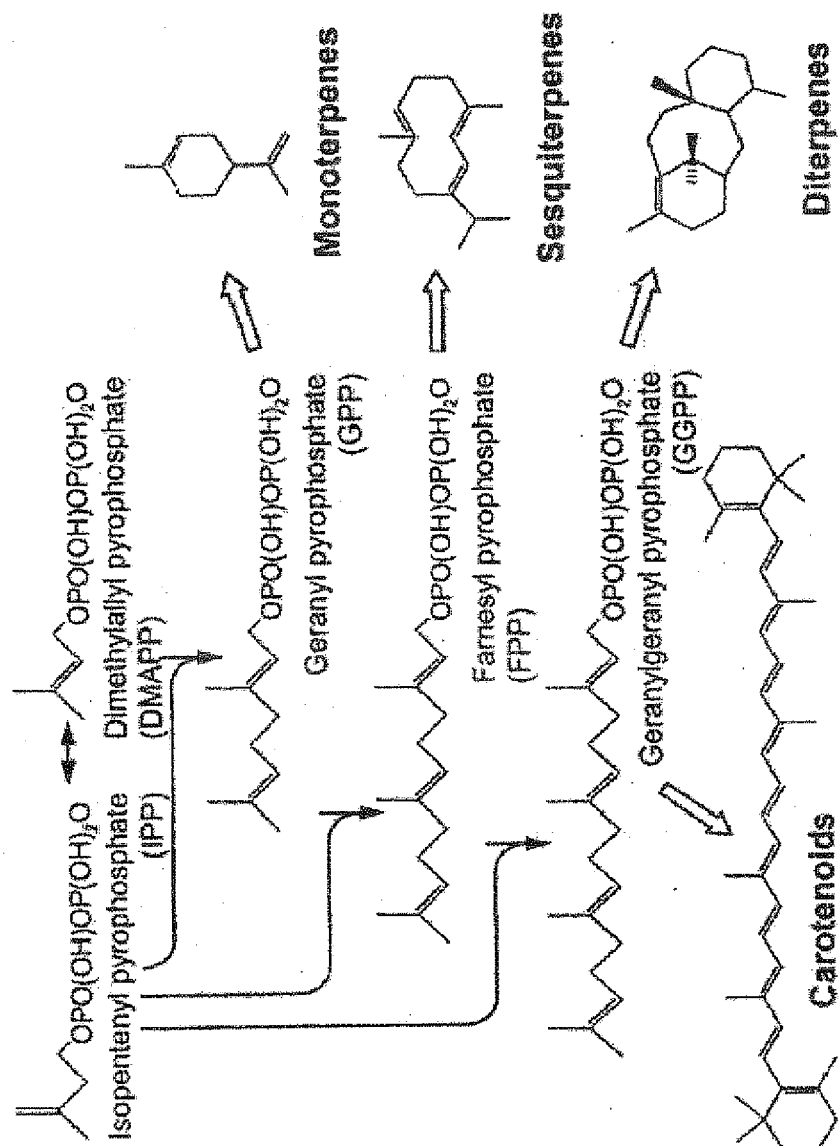


Figure 3

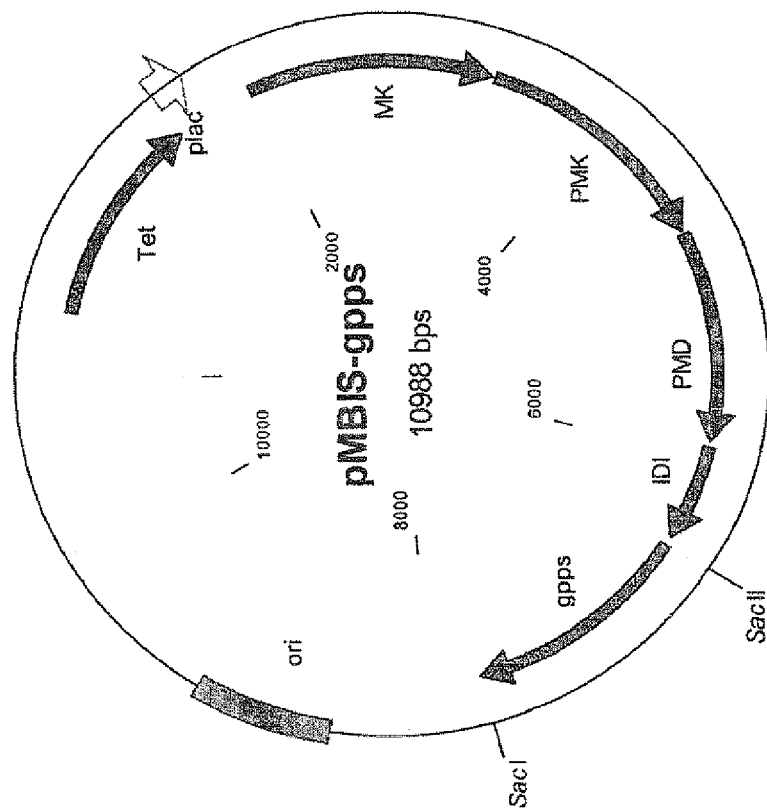


Figure 4

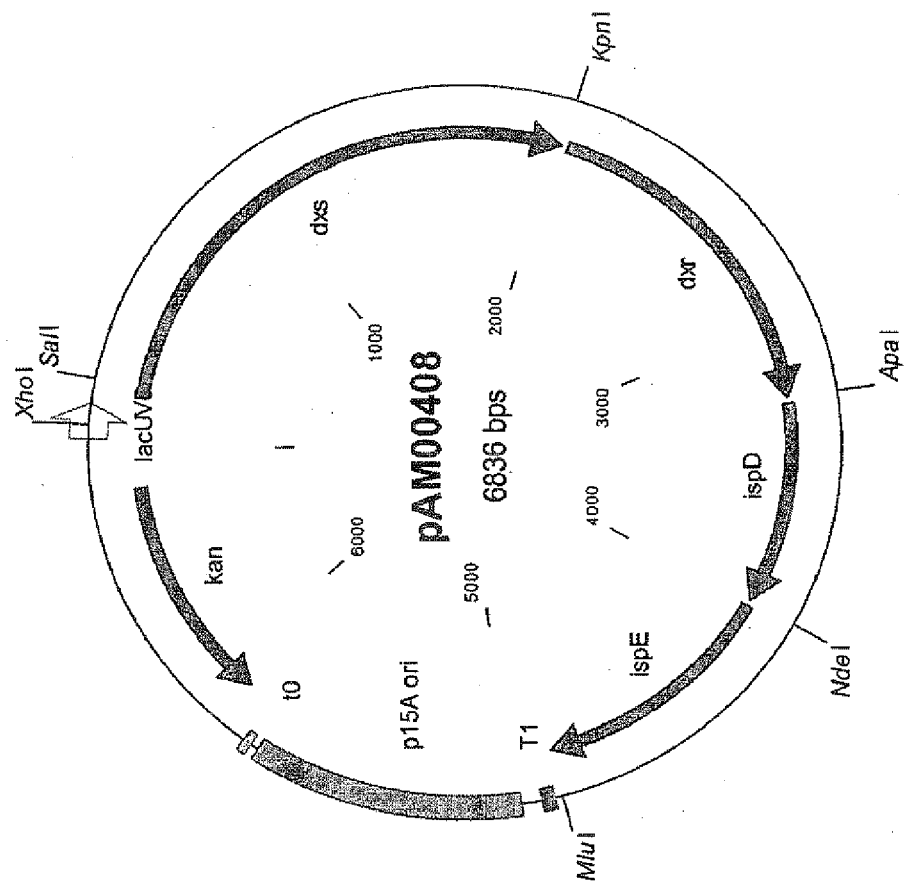


Figure 5

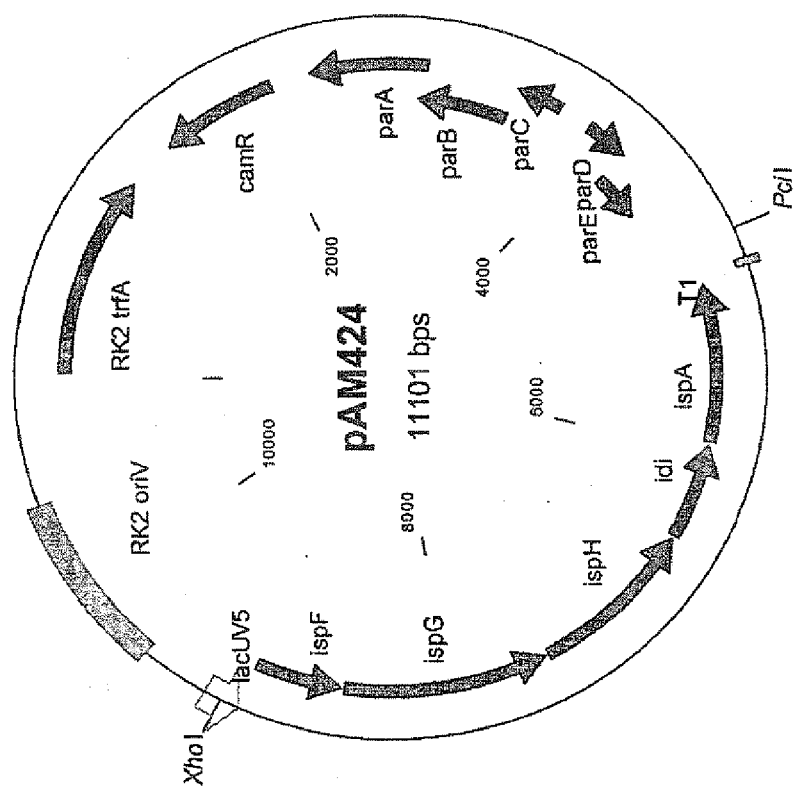
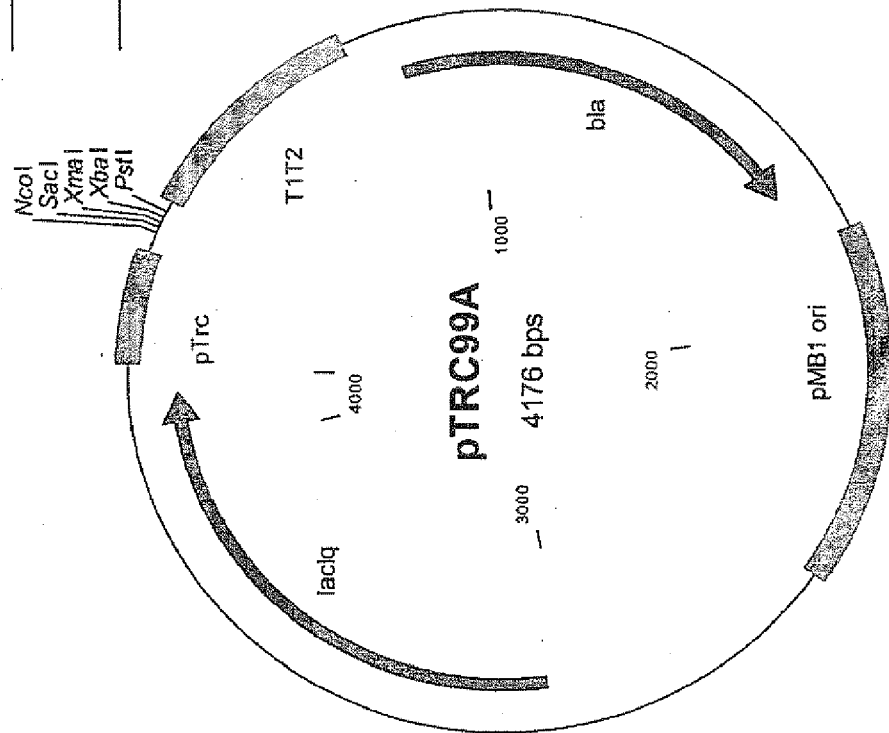


Figure 6



pTrc99A-ADS: ~1.6 kb ADS insert at *NcoI XmaI* site
 pTrc99A-FSA: ~1.7 kb FSA insert at *NcoI SacI* site
 pTrc99A-GTS: ~1.8 kb GTS insert at *XmaI XbaI* site
 pTrc99A-APS: ~1.9 kb APS insert at *XmaI XbaI* site
 pTrc99A-TS: ~1.9 kb TS insert at *XmaI XbaI* site
 pTrc99A-BPS: ~1.7 kb BPS insert at *XmaI XbaI* site
 pTrc99A-PHS: ~1.9 kb PHS insert at *XmaI XbaI* site
 pTrc99A-SS: ~1.8 kb SS insert at *XmaI XbaI* site
 pTrc99A-LLS: ~1.7 kb LLS insert at *NcoI XmaI* site
 pTrc99A-CS: ~1.8 kb CS insert at *NcoI XmaI* site
 pTrc99A-LMS: ~1.9 kb LMS insert at *XbaI PstI* site
 pAM373: ~1.7 kb FSA insert at *NcoI SacI* site

Figure 7

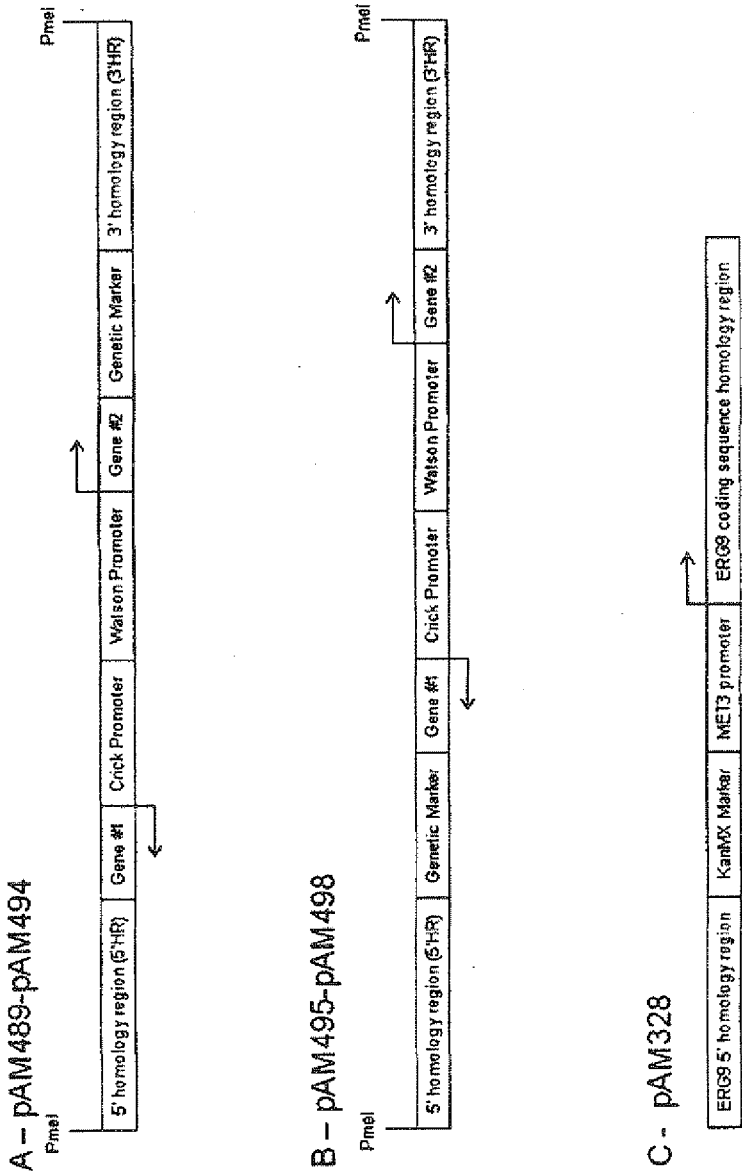


Figure 8

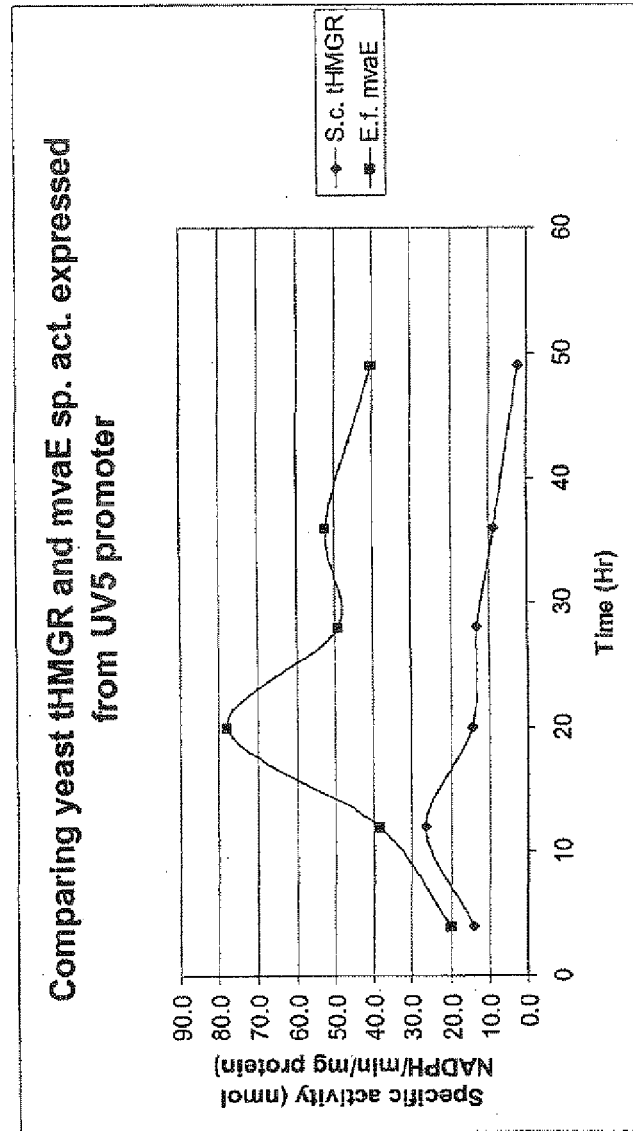


Figure 9

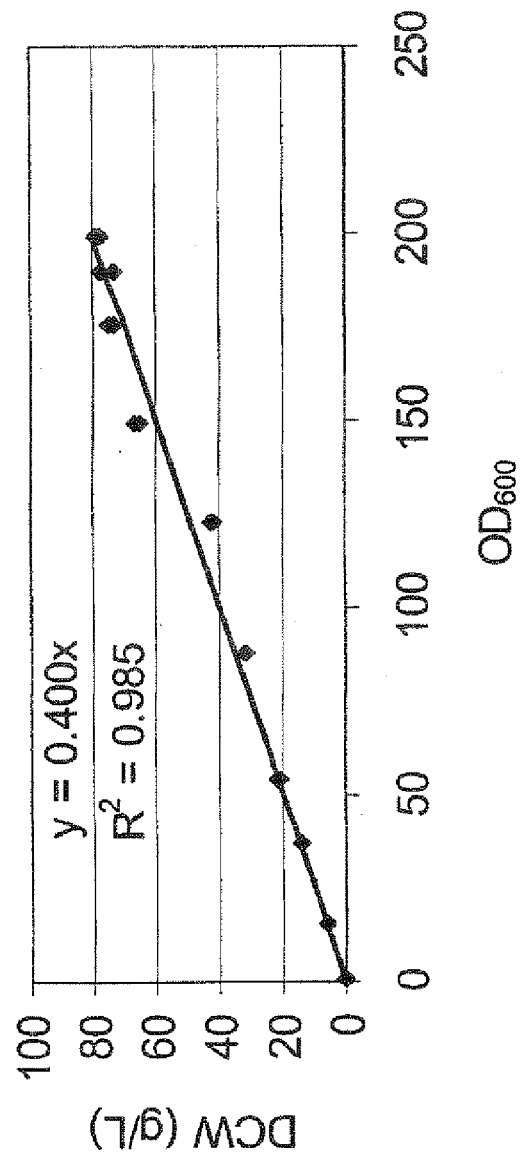


Figure 10A-B

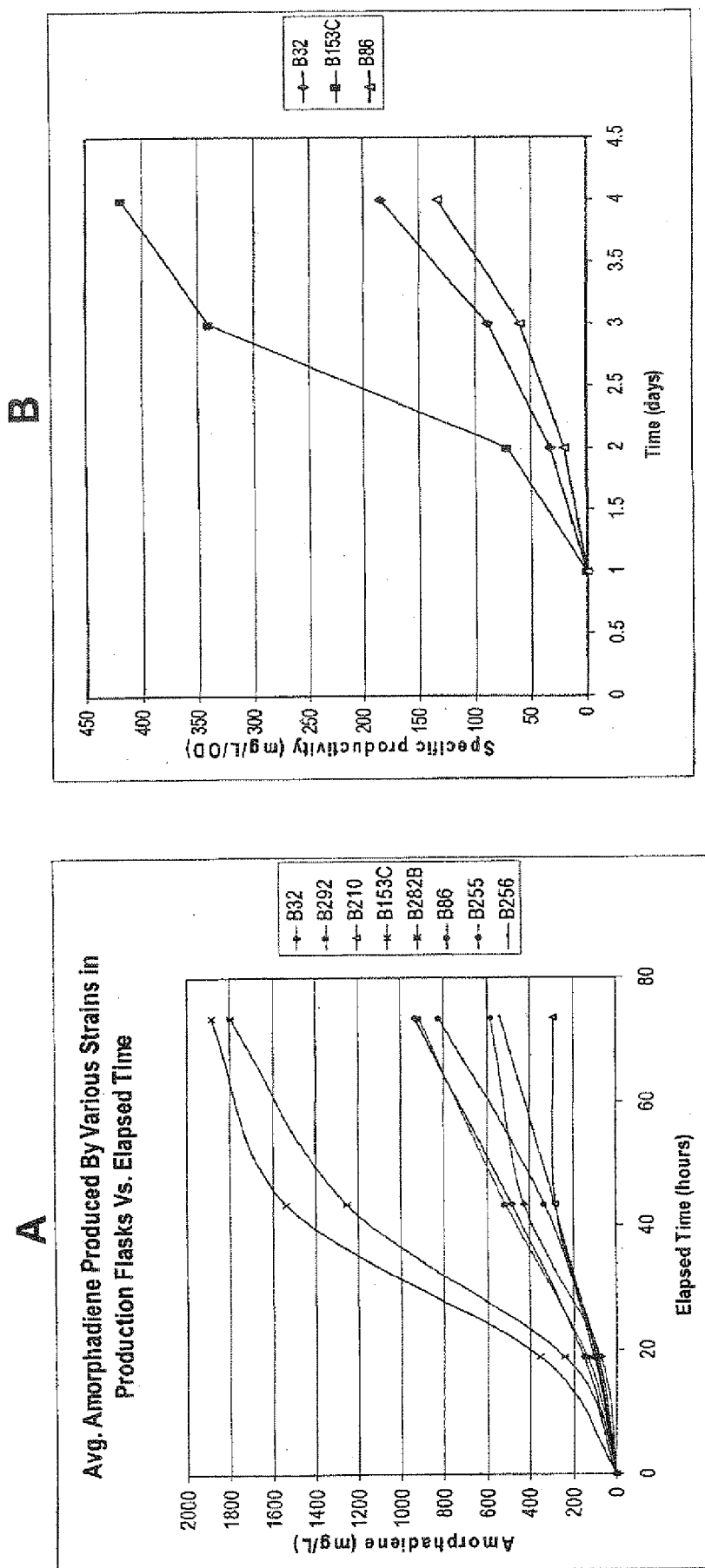


Figure 11A-B

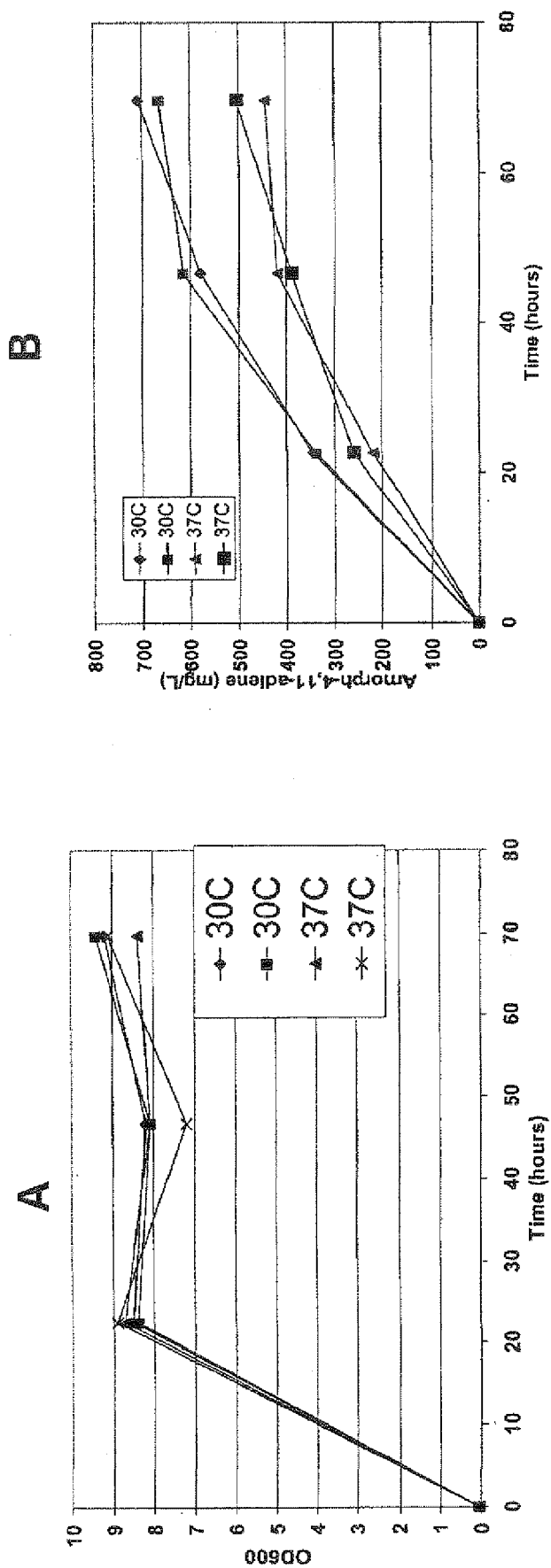


Figure 12A-B

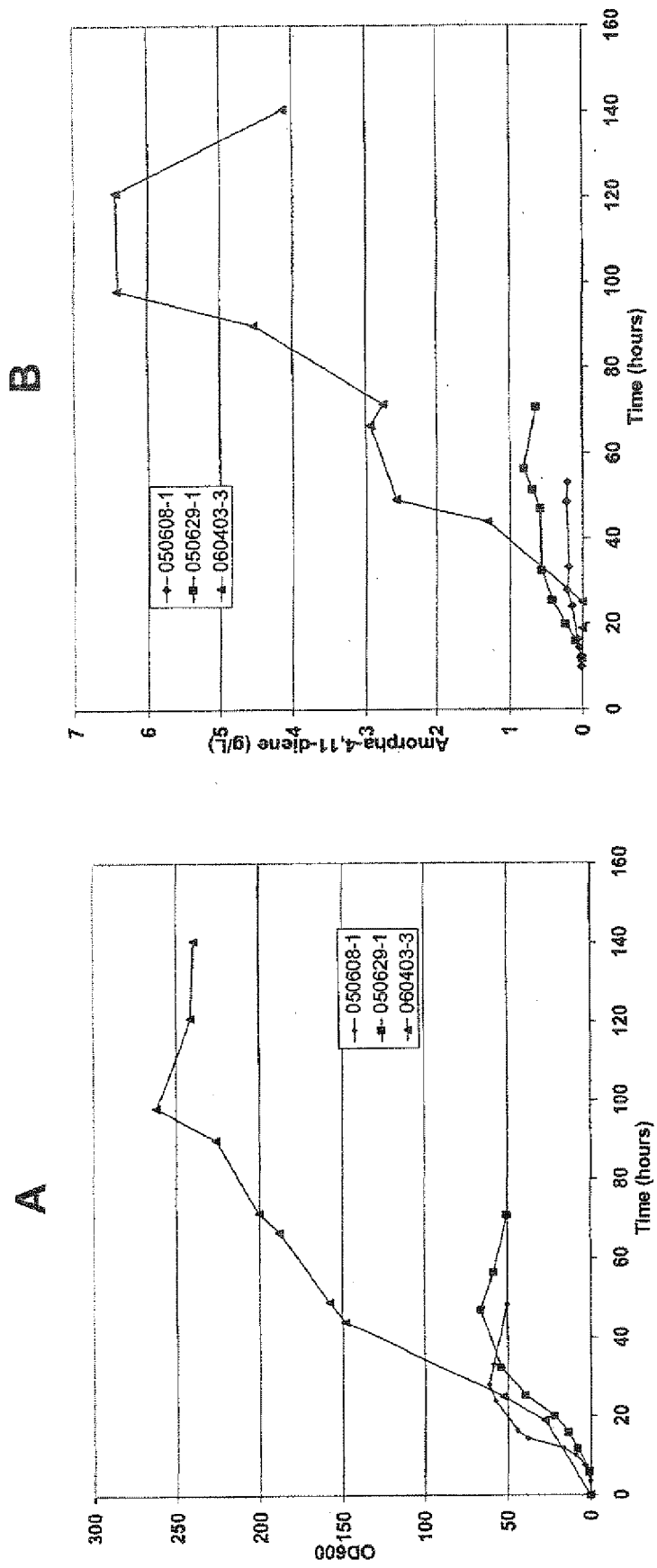


Figure 12C-D

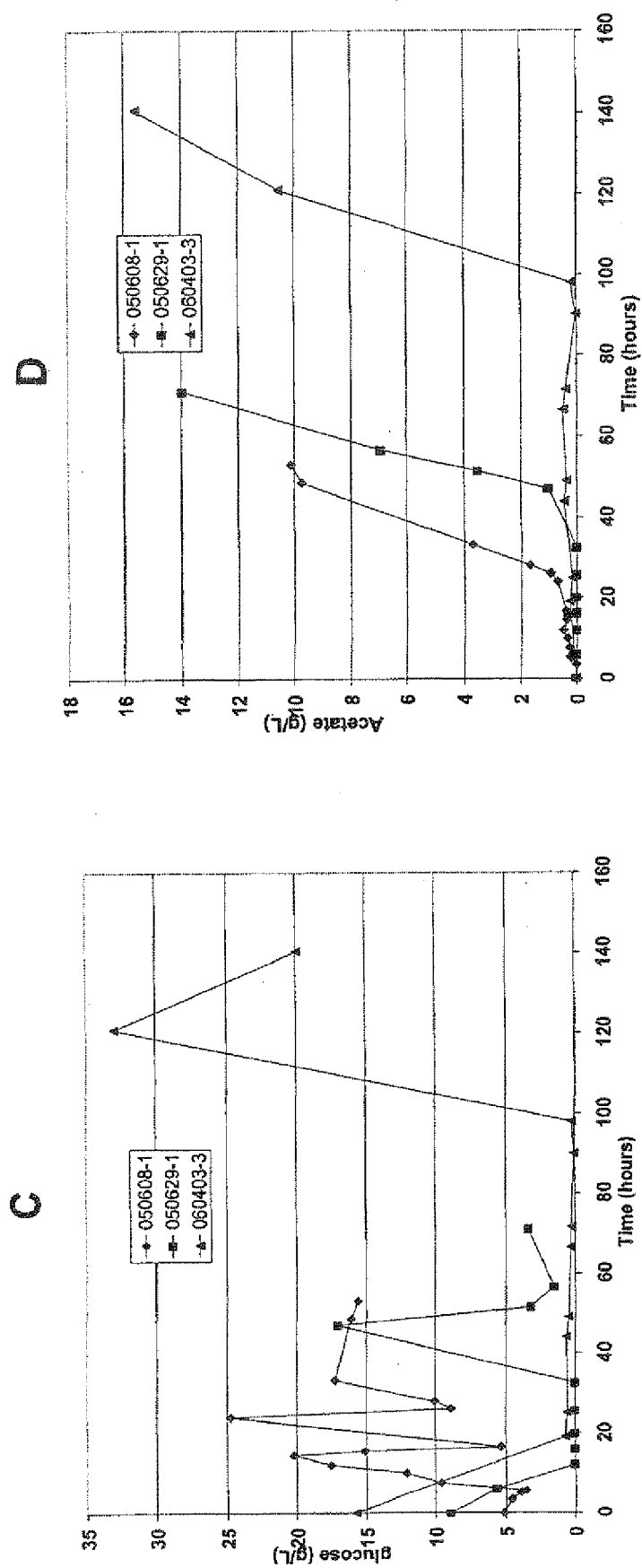


Figure 13A-B

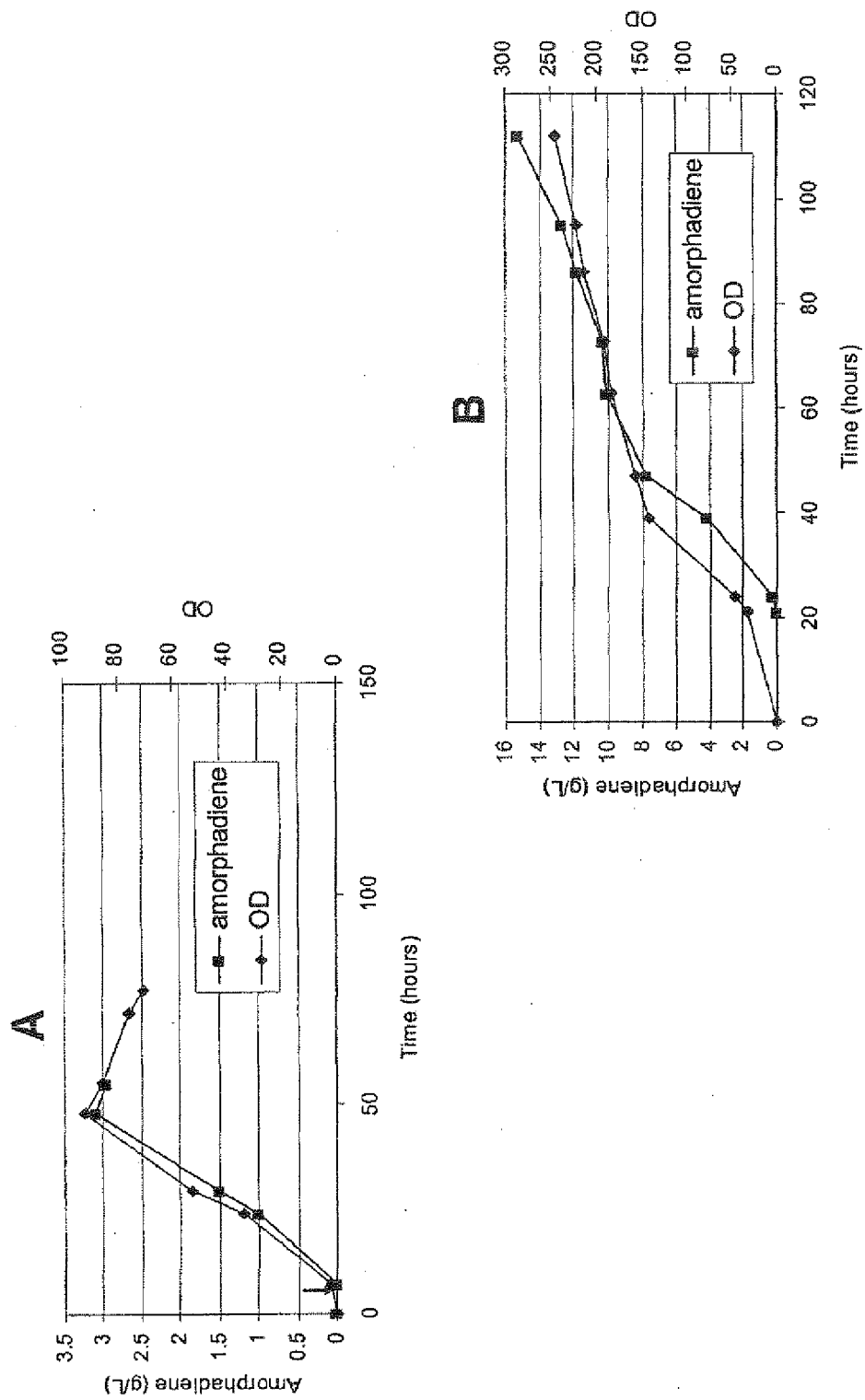


Figure 14A-B

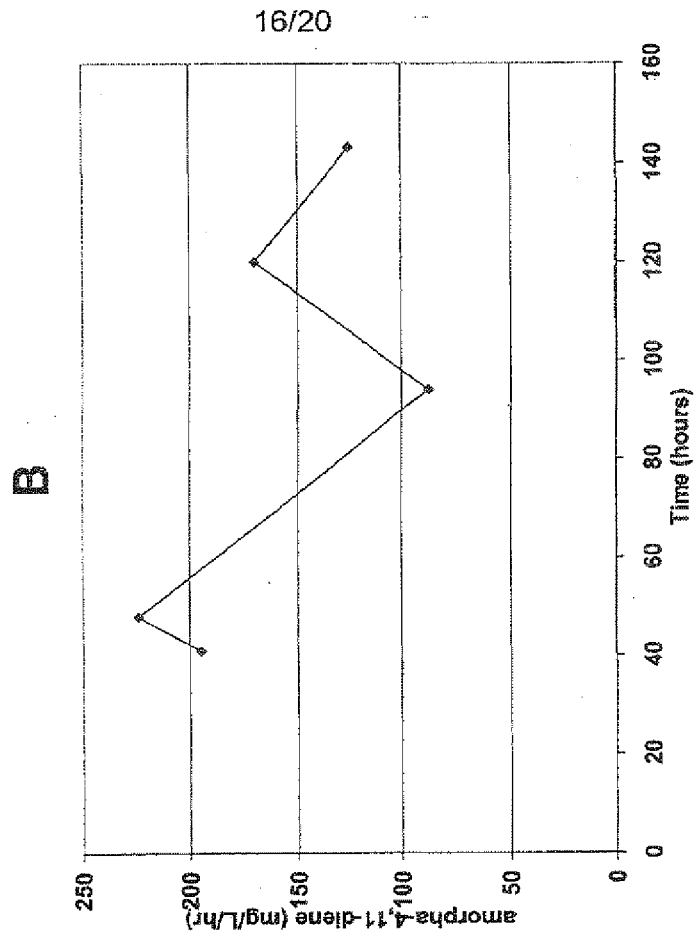
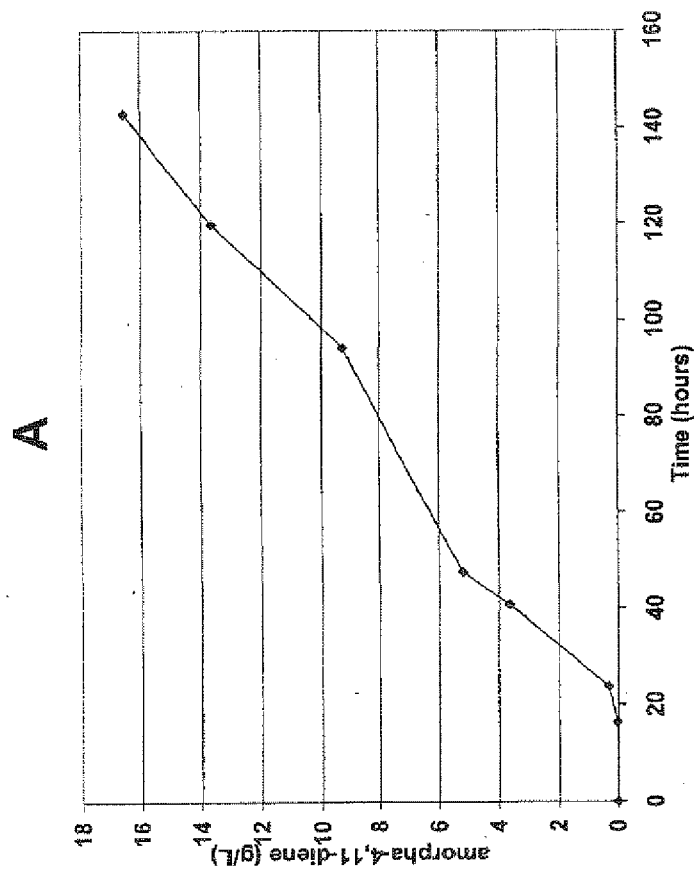


Figure 14C-E

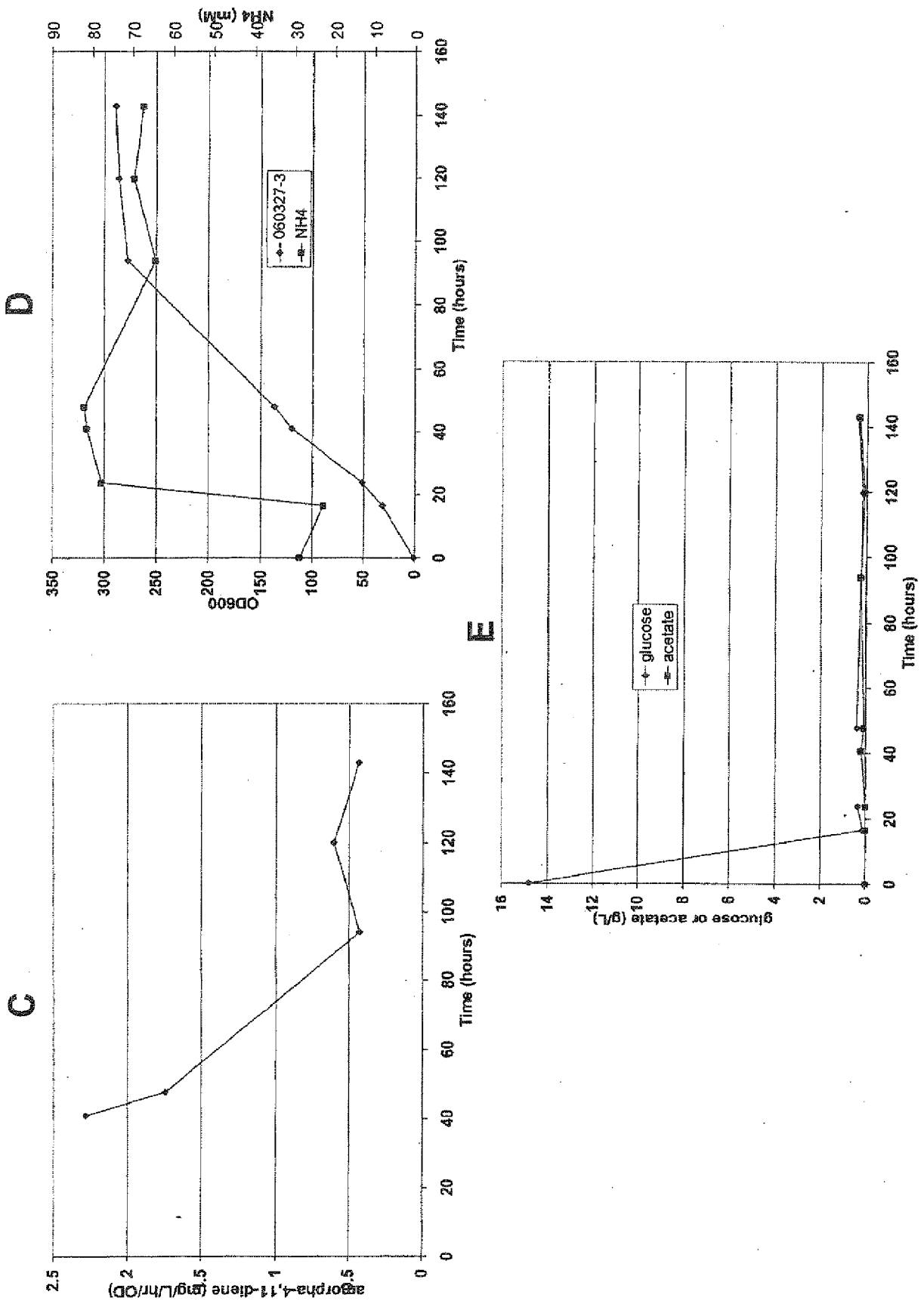


Figure 15A-B

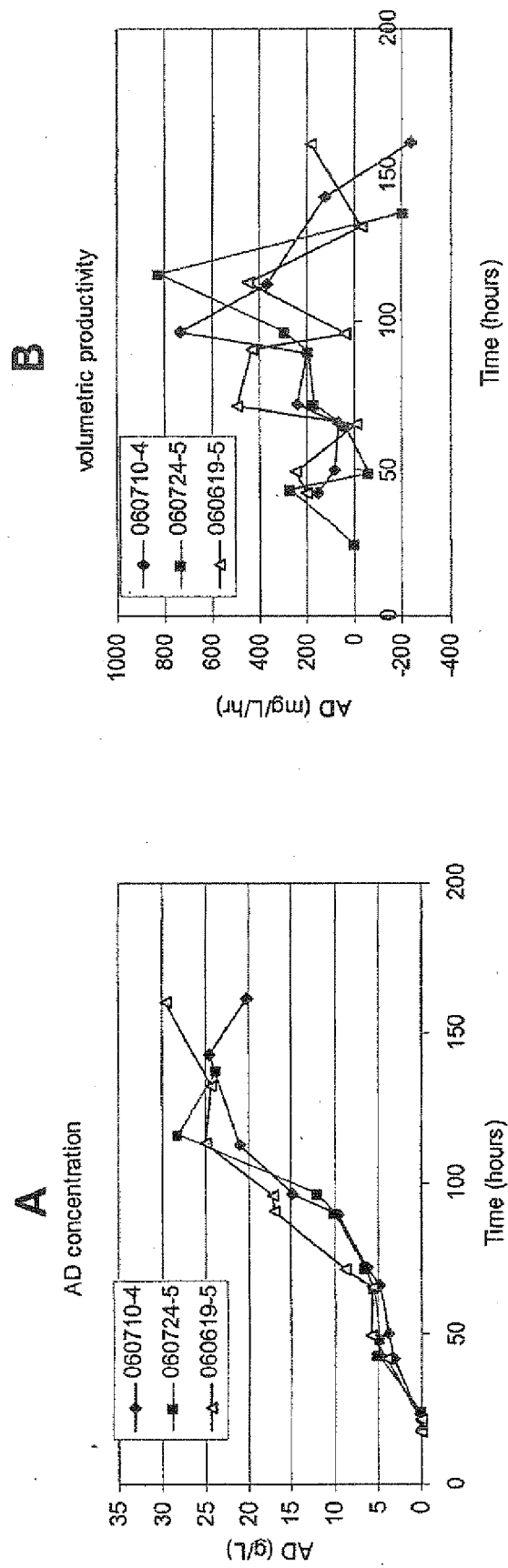


Figure 15C-E

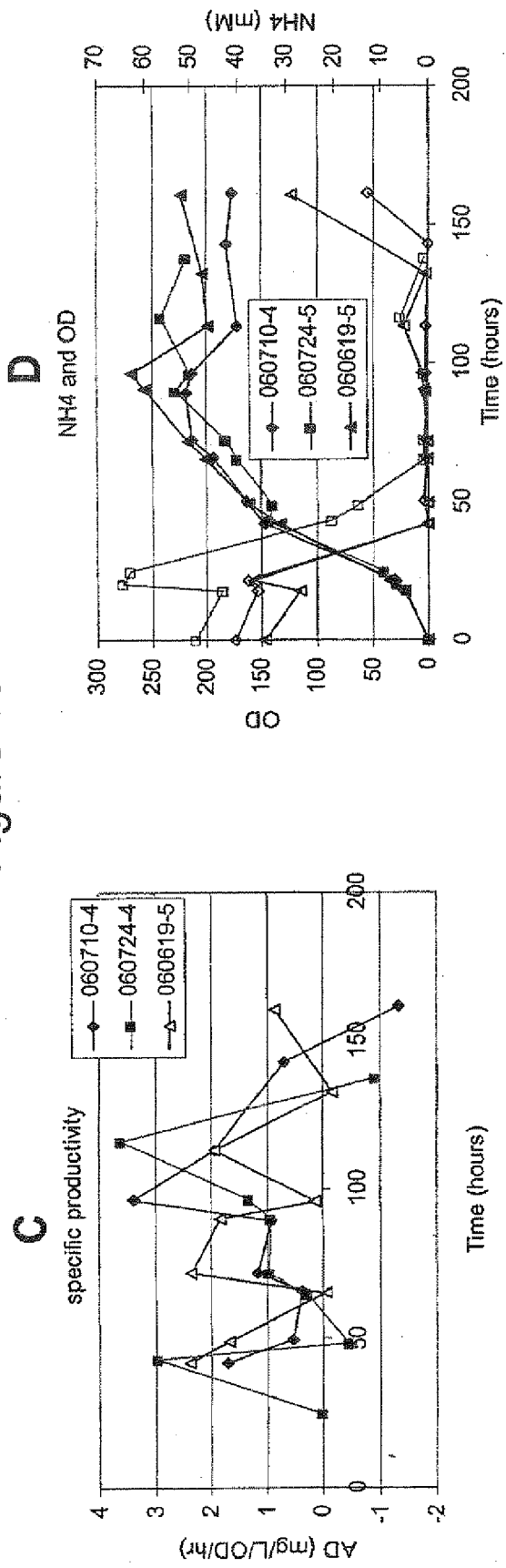
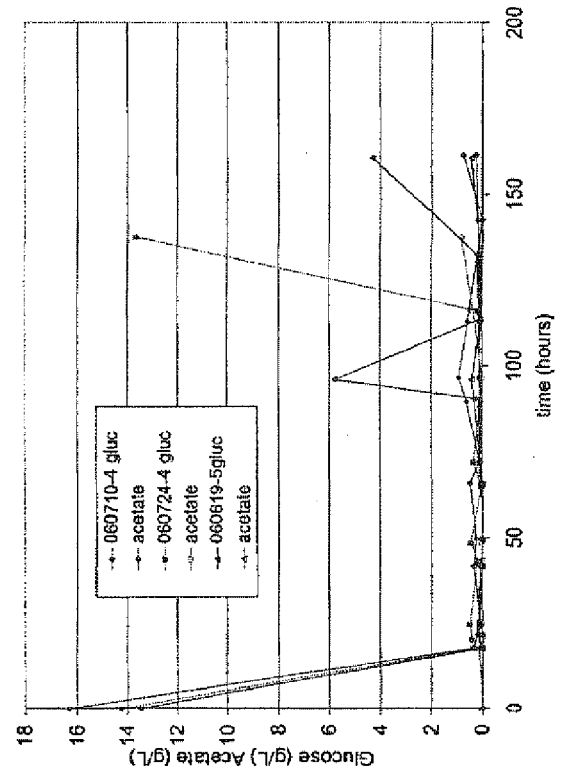
**E**
acetate and glucose

Figure 16

