



(12) **CORRECTED NEW EUROPEAN PATENT SPECIFICATION**

(15) Correction information:
Corrected version no 1 (W1 B2)
Corrections, see
Claims EN 2

(51) International Patent Classification (IPC):
C12P 7/02 (2006.01) **C12P 7/04** (2006.01)
C12P 5/02 (2006.01) **C12P 5/00** (2006.01)
C12M 1/02 (2006.01) **C10L 1/02** (2006.01)

(48) Corrigendum issued on:
20.04.2022 Bulletin 2022/16

(52) Cooperative Patent Classification (CPC):
C10L 1/02; C12M 21/12; C12M 27/02; C12M 27/20;
C12M 41/18; C12M 41/22; C12N 15/52;
C12P 5/007; C12P 7/02; C12P 7/04; Y02E 50/30

(45) Date of publication and mention
of the opposition decision:
26.01.2022 Bulletin 2022/04

(86) International application number:
PCT/US2007/012467

(45) Mention of the grant of the patent:
29.10.2014 Bulletin 2014/44

(87) International publication number:
WO 2007/139924 (06.12.2007 Gazette 2007/49)

(21) Application number: **07777278.8**

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

(54) **Bio-organic compound production system**

System zur Herstellung von bio-organischen Verbindungen

Système de fabrication de composés bio-organiques

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE
SI SK TR
Designated Extension States:
AL BA HR MK RS

(30) Priority: **26.05.2006 US 808989 P**
26.05.2006 US 808666 P
18.12.2006 US 870592 P
10.04.2007 US 922782 P

(43) Date of publication of application:
11.02.2009 Bulletin 2009/07

(73) Proprietor: **Amyris, Inc.**
Emeryville, CA 94608 (US)

(72) Inventor: **RENNINGER, Neil, Stephen**
Oakland, CA 94602 (US)

(74) Representative: **Mewburn Ellis LLP**
Aurora Building
Counterslip
Bristol BS1 6BX (GB)

(56) References cited:
WO-A-2005/078110 CH-A5- 677 791
DE-A1- 3 138 493 DE-B3-102005 018 256

US-A- 4 865 973 US-A- 5 763 237
US-A1- 2002 168 733 US-A1- 2004 229 326

- **ANTONIA ROJAS ET AL:**
"BIOTRANSFORMATION IN DOUBLE-PHASE
SYSTEMS: PHYSIOLOGICAL RESPONSES OF
PSEUDOMONAS PUTIDA DOT-T1E TO A
DOUBLE PHASE MADE OF ALIPHATIC
ALCOHOLS AND BIOSYNTHESIS OF
SUBSTITUTED CATECHOLS" APPLIED AND
ENVIRONMENTAL MICROBIOLOGY, vol. 70, no.
6, June 2004 (2004-06), pages 3637-3643,
XP002456441
- **STARK D ET AL: "In situ product removal (ISPR)**
in whole cell biotechnology during the last twenty
years", ADVANCES IN BIOCHEMICAL
ENGINEERING, BIOTECHNOLOGY, SPRINGER,
BERLIN, DE, vol. 80, 1 January 2003 (2003-01-01),
pages 149-175, XP009043803, ISSN: 0724-6145
- **SCHUGERL K: "Integrated processing of**
biotechnology products", BIOTECHNOLOGY
ADVANCES, ELSEVIER PUBLISHING, BARKING,
GB, vol. 18, no. 7, 1 November 2000 (2000-11-01),
pages 581-599, XP004317587, ISSN: 0734-9750,
DOI: 10.1016/S0734-9750(00)00051-3

- | | |
|---|--|
| <ul style="list-style-type: none"> • D. STARK ET AL: "Extractive Bioconversion of 2-Phenylethanol from L-Phenylalanine by <i>Saccharomyces cerevisiae</i>", BIOTECHNOLOGY PROGRESS, vol. 18, no. 3, 7 June 2002 (2002-06-07), pages 514-523, XP055049324, ISSN: 8756-7938, DOI: 10.1021/bp020006n • BYUN S Y ET AL: "Two-phase airlift fermentor operation with elicitation for the enhanced production of benzophenanthridine alkaloids in cell suspensions of <i>Escherichia californica</i>", BIOTECHNOLOGY AND BIOENGINEERING, WILEY & SONS, HOBOKEN, NJ, US, vol. 44, no. 1, 5 June 1994 (1994-06-05), pages 14-20, XP002032586, ISSN: 0006-3592, DOI: 10.1002/BIT.260440104 | <ul style="list-style-type: none"> • DATABASE WPI Week 200633 Thomson Scientific, London, GB; AN 2006-311267 & JP 2006 109784 A (MITSUBISHI GAS CHEM CO INC) 27 April 2006 (2006-04-27) • NEWMAN J D ET AL: "High-level production of amorpho-4,11-diene in a two phase partitioning bioreactor of metabolically engineered <i>Escherichia coli</i>", BIOTECHNOLOGY AND BIOENGINEERING, WILEY & SONS, HOBOKEN, NJ, US, vol. 95, 28 July 2006 (2006-07-28), pages 684-691, XP002592776, ISSN: 0006-3592, DOI: 10.1002/BIT.21017 |
|---|--|

Description**BACKGROUND OF THE INVENTION**

[0001] Traditionally, bio-organic compounds of interest have been manufactured by extraction from natural sources such as plants, microbes, and animals. However, extraction yields are usually very low as most bio-organic compounds accumulate in nature in small amounts. Given that these quantities are far less than is for many commercial applications, there remains a need for systems and procedures that produce bio-organic compounds on an industrial scale. Newman et al, Biotechnology and Bioengineering, vol. 95, 28 July 2006, pages 684-691 describe high-level production of amorpho-

[0002] The present invention addresses this need. Provided are various industrial-scale systems for making bio-organic compounds using host cells. These bio-organic compounds have at least five carbon atoms and can be a carbohydrate such as a mono- or poly-alcohol, ester, ether, aldehyde, ketone, or a hydrocarbon such as an alkane, alkene, or alkyne. The bio-organic compound can be linear or cyclic and can be saturated or unsaturated.

SUMMARY OF THE INVENTION

[0003] The present invention provides various bio-organic compound production systems in accordance with the claims. In one aspect, a bio-organic compound production system is provided which comprises:

- a. at least one vessel having a capacity of at least 100 liters;
- b. an aqueous medium, within the vessel, comprising a first phase;
- c. a plurality of genetically modified bacterial or fungal host cells, within the aqueous medium, capable of converting a carbohydrate carbon source present in the aqueous medium into at least one bio-organic compound having at least ten carbon atoms; and
- d. a liquid organic second phase, comprising the at least one bio-organic compound in an amount of at least 90%, in contact with the first phase; wherein the at least one bio-organic compound is an isoprenoid compound; and wherein the organic second phase is spontaneously formed.

[0004] Also disclosed herein is a method of producing at least one bio-organic compound is provided. The method comprises:

- a. culturing, in an aqueous medium, a plurality of genetically modified bacterial or fungal host cells that produce at least one bio-organic compound having at least ten carbon atoms, wherein the aqueous medium forms a first phase, and wherein cells in the plurality convert a carbohydrate carbon source present in the aqueous medium into the at least one bio-organic compound having at least ten carbon atoms;
- b. forming a liquid organic second phase, comprising the at least one bio-organic compound in an amount of 40-99%, in contact with the first phase;
- c. separating at least a portion of the second phase from the first phase; and
- d. isolating the at least one bio-organic compound from the second phase; wherein the at least one bio-organic compound is an isoprenoid compound; and wherein the organic second phase is spontaneously formed.

BRIEF DESCRIPTION OF THE DRAWINGS**[0005]**

Figure 1 is a vessel having a capacity of at least 100 liters for use in the present invention.

Figure 2 is another vessel embodiment.

Figure 3 is a schematic representation of the mevalonate ("MEV") pathway for the production of isopentenyl diphosphate ("IPP").

Figure 4 is a schematic representation of the DXP pathway for the production of 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.

Figure 5 is a schematic representation of the conversion of IPP and DMAPP to geranyl pyrophosphate ("GPP"),

farnesyl pyrophosphate ("FPP"), and geranylgeranyl pyrophosphate ("GGPP").

Figure 6 shows a map of expression plasmid pMBIS-gpps.

Figure 7 shows a map of expression plasmid Pam00408

Figure 8 shows a map of expression plasmid pAM424.

Figure 9 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.

Figure 10 are schematics for the construction of plasmids pAM489-pAM498.

DETAILED DESCRIPTION OF THE INVENTION

Definition

[0006] 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:

[0007] "Bio-organic compound" refers to an organic compound having at least five carbon atoms that can be made by a host cell by taking a carbohydrate carbon source and converting the carbohydrate carbon source into the desired product.

[0008] "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 4.

[0009] "Endogenous" refers to a substance or process that can occur naturally, e.g., in a non-recombinant host cell.

[0010] "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).

[0011] "Host cell" and "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.

[0012] "Isoprenoid" and "isoprenoid compound" are used interchangeably herein and refer to a compound derivable from isopentenyl diphosphate.

[0013] "Isolate" and "isolating" when referred to a bio-organic compound is the enrichment of the amount of the bio-organic compound in a composition. Consequently, the amount of the bio-organic compound in a composition after the bio-organic compound has been isolated or subject to an isolating step is greater than the amount present in the composition prior to such step.

[0014] "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 3.

[0015] "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.

[0016] "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.

[0017] "Pyrophosphate" is used interchangeably herein with "diphosphate".

[0018] As used herein, a composition that is a "substantially pure" compound is substantially free of one or more other compounds, i.e., the composition contains greater than 80 vol.%, greater than 90 vol.%, greater than 95 vol.%, greater than 96 vol.%, greater than 97 vol.%, greater than 98 vol.%, greater than 99 vol.%, greater than 99.5 vol.%, greater than 99.6 vol.%, greater than 99.7 vol.%, greater than 99.8 vol.%, greater than 99.9 vol.% of the compound; or less than 20 vol.%, less than 10 vol.%, less than 5 vol.%, less than 4 vol.%, less than 3 vol.%, less than 2 vol.%, less than 1 vol.%, less than 0.5 vol.%, less than 0.1 vol.%, or less than 0.01 vol.% of the one or more other compounds, based on the total volume of the composition.

[0019] In the following description, all numbers disclosed herein are approximate values, regardless whether the word "about" or "approximate" is used in connection therewith. They may vary by 1 percent, 2 percent, 5 percent, or, sometimes, 10 to 20 percent. Whenever a numerical range with a lower limit, RL and an upper limit, RU, is disclosed, any number falling within the range is specifically disclosed. In particular, the following numbers within the range are specifically disclosed: $R=RL+k*(RU-RL)$, wherein k is a variable ranging from 1 percent to 100 percent with a 1 percent increment, i.e., k is 1 percent, 2 percent, 3 percent, 4 percent, 5 percent,..., 50 percent, 51 percent, 52 percent,..., 95 percent, 96 percent, 97 percent, 98 percent, 99 percent, or 100 percent. Moreover, any numerical range defined by two R numbers as defined in the above is also specifically disclosed.

[0020] In addition to the definitions above, certain compounds described herein have one or more double bonds that can exist as either the Z or E isomer. The invention in certain embodiments encompasses these compounds as individual isomers in a substantially pure form as well as mixtures of various isomers, e.g., racemic mixtures of stereoisomer.

Apparatus for Making Bio-organic Compounds

[0021] The present invention provides various production systems for making bio-organic compounds. In some embodiments, the bio-organic compounds may be produced using batch, continuous, fed-batch or semi-continuous fermentation processes.

[0022] Batch fermentation may be a closed system where the composition of the media is fixed at the beginning of the fermentation and not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the media is inoculated with the desired organism or organisms and fermentation is permitted to occur adding nothing to the system. In some embodiments, however, "batch" fermentation is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems the metabolite and biomass compositions of the system may change constantly up to the time the fermentation is stopped. Within batch cultures, cells may moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

[0023] A variation on the standard batch system is the fed-batch system. Fed-batch fermentation processes are also suitable in the present invention and comprise a typical batch system with the exception that additional carbon source or substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in fed-batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂.

[0024] Continuous fermentation is an open system where a defined fermentation media is added continuously to one or more bioreactors which may be in series and an equal amount of conditioned media is removed simultaneously from the system for additional processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in the fermentation.

[0025] Accordingly, in some embodiments of the invention, a bio-organic production system is provided which comprises:

- a. a vessel having a capacity of at least 100 liters;
- b. an aqueous medium, within the vessel, forming a first phase;
- c. a plurality of genetically modified bacterial or fungal host cells, within the aqueous medium, capable of converting a carbohydrate carbon source present in the aqueous medium into at least one bio-organic compound having at least ten carbon atoms; and
- d. a liquid organic second phase, comprising the at least one bio-organic compound in an amount of at least 90%, in contact with the first phase;

wherein the at least one bio-organic compound is an isoprenoid compound; and wherein the organic second phase is spontaneously formed.

[0026] A suitable vessel for use in the present invention can be any vessel for holding the host cells and aqueous medium for fermentation. For example, the vessel can be a tank for a reactor or fermenter or it can be a part of a centrifuge that can separate heavier materials from lighter materials in subsequent processing steps. Alternatively, one or a plurality

of vessels may be used in a continuous or semi-continuous process.

[0027] A general illustrative example of a suitable vessel 100 is shown in Figure 1. The vessel 100 includes: an inlet port 120 for the addition of host cells, fermentation media, and other compounds, nutrients or compositions to assist, regulate or improve fermentation of the host cells, production of the bio-organic compound or compounds, and performance of additional production steps into the vessel; an outlet port 130 for removing the materials during or at the end of the fermentation process, and a gas outlet 140 for venting off exhaust gases such as carbon dioxide produced during or after the fermentation process. Vessel 100 may be completely filled with host cells, fermentation media and other materials so that there is no space for gas at the top of the vessel. Alternatively, vessel 100 can be partially filled thus leaving void space occupied by a gas. The amount, pressure and composition of the gas in the void space may be controlled to optimize or maximize growth of the host cells and production of the bio-organic compound or bio-organic compounds. For example, during fermentation of aerobic host cells, the gas typically may comprise air or other oxygen-containing gas at various pressures above, at or below atmospheric pressure, for example for microaerophilic and anaerobic host cells the oxygen concentration of the gas may be controlled within a range lower than atmospheric concentration while still above zero while during fermentation for anaerobic host cells, the gas typically has little to no oxygen and can completely comprise mostly or completely of nitrogen or other suitable gas.

[0028] In a closed system, inlet port 120, outlet port 130 and gas outlet 140 of vessel 100 shown in Figure 1 may be closed or under positive pressure during the fermentation process. Alternatively, particularly when using aerobic host cells, vessel 100 can be used as an open system whereby one or more of the ports and outlet are opened to the atmosphere providing a system for gas/liquid mass transfer (air or oxygen in and carbon dioxide out). If desired, gas outlet 140 may function both as a gas outlet and as a gas inlet where oxygen or air or other gas may be introduced into the system. In some embodiments, vessel 100 includes separate gas inlets and separate gas outlets. In such open systems, additional hardware may be included on the vessel for preventing contamination or infiltration of other organisms or other materials into the vessel during the fermentation.

[0029] Another vessel embodiment is illustrated in Figure 2. In addition to inlet port 220, outlet port 230, gas inlet 235, and gas outlet 240 similar to the vessel in Figure 1, the vessel 200 of Figure 2 includes an agitator 250 for mixing. In some embodiments, agitator 250 may comprise a motor-driven shaft 252 which may include a shaft seal 251 and is connected to one or more impellers 254. Agitator 250 may be typically attached to the top or bottom of the vessel 200. Optionally, each impeller 254 may be terminated with one or more paddles 256. Impellers 254 may be any suitable shape and may be selected specifically to control amount of mixing, growth rate of the host cells, production rate of the bio-organic compound, shear rate and oxygen or other gas transfer rates. Additionally, one or more baffles 258 can be added to the vessel 200 to further improve mixing. In another embodiment, agitation may be supplied in the form of a recycle line with a pump that draws material from one portion of the vessel such as the bottom and reintroduces the material into the vessel at another portion of the vessel such as the top. Agitation within the vessel of the host cells and the fermentation medium aids in ensuring that the host cells are exposed to adequate nutrients to enable them to grow and produce the bio-organic compounds.

[0030] If the fermentation process is an aerobic process, oxygen or air can be bubbled through a sparger 260 for improved gas/liquid mass transfer. The sparger 260 may include one or more gas outlets (not shown) that are submerged within the fermentation media, preferably at or near the bottom of the vessel. In some embodiments, the sparger 260 may be a sparging ring having multiple gas outlets arranged in a generally circular or round configuration. Alternatively, for shear sensitive organisms or to reduce foaming, passive aeration of the vessel may be provided, such as use of various aeration screens, membranes, fibers or other passive aeration devices or by removing a portion of the media from the vessel, oxygenating it and returning it to the vessel.

[0031] If temperature control is desired, then a heater or heat exchanger may be used to heat or cool the fermentation reaction. In one embodiment, the temperature may be controlled using a heating/cooling jacket 270 surrounding and/or attached to at least a portion of vessel 200 that may be connected to a heat exchanger (not shown) that circulates temperature controlled heat exchange fluid through jacket 270. Alternatively, a heater, or heat exchanger may be immersed in the fermentation medium. Illustrative examples of this type of heater or heat exchanger include an electric immersion heater, an immersed coiled or linear tube heat exchanger carrying a heat-exchange fluid such as heated water or oil, and one or more spargers that inject a heated stream such as air and/or water into the fermentation medium. Alternatively or additionally, a heater or heat exchanger can be attached to the outside of the vessel. Such heaters and heat exchangers include electrical heat tape on outside sidewalls of the vessel and heated or jacketed recycle lines attached to the vessel.

[0032] Vessel 200 can include additional inlet and outlet ports. In some embodiments, the additional inlet and outlet ports may be located on the top, sides or bottom of the vessel 200. In some embodiments, the additional inlet ports include feed lines for the addition of oxygen or other gases, nutrients, foam and pH control agents during the fermentation reaction. Any of the inlet and outlet ports may include sterilization mechanisms for multiple uses including in-process use, and multiple connection or reconnection during the fermentation process.

[0033] In addition, one or more probe ports 280 and/or sampling valves 290 can be positioned at various places on

vessel 200 to help monitor critical parameters such as concentrations of various products and metabolites, pH, liquid level, pressure, foam, dissolved oxygen concentration, temperature, agitation rate, power, voltage, valve positions and cell density during the fermentation process.

[0034] It should be understood that the vessels in Figures 1 and 2 are for illustrative purposes and that many different vessel configurations for the fermentation process may be used, for example, according to the type of host cell, the bio-organic compound or compounds produced, the production volume, the type of fermentation process, the type of downstream processing, the separation process and other considerations.

[0035] A vessel such as that shown in Figure 2 is suitable for use in batch fermentation processes. If a continuous or semi-continuous fermentation process is desired (as opposed to a batch fermentation process) where materials are constantly added to or withdrawn from the vessel, the vessel typically includes additional inlet and outlet ports which may be located on the top, bottom or on the sides of the vessel. These additional inlet and outlet ports facilitate the flow of materials in and out of the vessel. In some embodiments, one or more vessels continuously receive host cells, fermentation medium, and optional additives while continuously discharging host cells, byproducts, and/or bio-organic compounds from the vessels. In these embodiments, the discharge from one vessel may be used as the feedstock to another vessel that optionally also receives fresh host cells, fermentation medium, nutrients, and/or other additives. A single vessel or a series of vessels together can be configured to provide the desired average residence time for the host cells. A portion of the discharge from one of the down-stream vessels can be returned to one or more upstream vessels to recycle the discharge to an earlier stage of processing, or other materials from processing steps further downstream can be reintroduced into the vessels.

[0036] The vessels used in some embodiments of the present invention include additional hardware that may be attached to the vessel to facilitate processing. Such hardware may include additional hardware for facilitating clean-in-place and sterilize-in-place processing. In some embodiments, one, some or each of the ports, outlets, inlets, valves and all of the hardware inside the vessel may be sterilized in place. In some embodiments, the sterilization may occur using steam sterilization. For example, any of the ports, outlets or sampling valves may include or have attached to them additional hardware that provides for steam supply to and condensate return from the port outlet or valve such that it may be steam sterilized prior to use or reuse.

[0037] The vessel or vessels may have a capacity of at least 100 liters. In some embodiments, the vessel has a capacity of from 100 to 3,000,000 liters such as at least 1000 liters, at least 5,000 liters, at least 10,000 liters, vessel at least 25,000 liters, at least 50,000 liters, at least 75,000 liters, at least 100,000 liters, at least 250,000 liters, at least 500,000 liters or at least 1,000,000 liters.

[0038] The vessel or vessels may include or have attached to them sensors and probes for measuring various parameters such as pressure, pH, dissolved oxygen concentration, temperature, gas flow rates, liquid flow rates, liquid level, valve positions, foaming, agitation, power, voltage and any other parameters useful in controlling or optimizing the growth of the host cells and the production of the bio-organic compound or compounds. The sensors and probes may feed information to one or more automated systems for controlling and recording the various parameters measured and for adjusting any of the various parameters by controlling air flowrates, power, heating or cooling to control vessel temperature, stirring rpms, pumps, sterilization or clean in place of the vessel or any of the inlet, outlet, addition, sampling valves or other ports, outlet flow control or any other relevant mechanism for controlling a parameter or parameters of the fermentation. Such adjustments may occur using any known control mechanism, such as for example, control or actuation of various valves, pumps or motors and may use proportional, proportional-integral or proportional -integral-derivative control systems.

[0039] The automated system or systems may additionally be controlled and monitored by a central control system, which may be a local or plant wide control system and may control production of just one bio-organic compound production process or multiple bio-organic compound production processes. The automated system or systems and central control system may comprise any suitable software, firmware and/or hardware, which may be proprietary or off the shelf or a combination thereof and may communicate using any suitable communication system. Non-limiting examples of such communication systems include hardwired systems that may be digital or analog, and may include direct connection or be in the form of a network such as a LAN or a WAN or ethernet. In addition, in some embodiments the communication system may be wireless and may be proprietary, BLUETOOTH, ultra wide band, 802.11 a,b,g or n or ZigBee, including TDMA, FDMA, OFDM, and CDMA and may operate in any suitable frequency band such as 2.4 GHz or 5.8 GHz.

[0040] Any of the vessels used in the production of the bio-organic compounds may include additional hardware, such as additional agitators, additional inlet ports, outlet ports, sampling ports, additional heating/cooling equipment, such as additional heating coils, additional aeration equipment such as additional spargers, additional sensors and probes, additional cleaning or sterilization equipment to facilitate processing or any other parameter of the fermentation.

[0041] Disclosed herein is a system for making C₁₀-C₂₀ isoprenoid compounds comprising:

- a. a vessel having a capacity of at least 100 liters;
- b. an aqueous medium, within the vessel, forming a first phase;

- c. a plurality of genetically modified bacterial or fungal host cells, within the aqueous medium, capable of converting a carbohydrate carbon source present in the aqueous medium into at least one C₁₀-C₂₀ isoprenoid; and
- d. a liquid organic second phase, comprising the at least one isoprenoid in an amount of 40-99%, in contact with the first phase; wherein the organic second phase is spontaneously formed.

[0042] In some embodiments, the isoprenoid compound or compounds is a C₁₀ isoprenoid. These compounds are derived from two isoprene units and are also called monoterpenes. An illustrative example of a monoterpene is myrcene. In other embodiments, the isoprenoid compound or compounds is a C₁₅ isoprenoid. These compounds are derived from three isoprene units and are also called sesquiterpenes. An illustrative example of a sesquiterpene is patchoulol (which is also known as patchouli alcohol). In other embodiments, the isoprenoid compound or compounds is a C₂₀ isoprenoid. These compounds are derived from four isoprene units and also called diterpenes. An illustrative example of a diterpene is taxadiene. In yet other examples, the isoprenoid compound or compounds is a C₂₀₊ isoprenoid. These compounds are derived from more than four isoprene units and include: triterpenes (C₃₀ isoprenoid compounds derived from 6 isoprene units) such as squalene; tetraterpenes (C₄₀ isoprenoid compounds derived from 8 isoprenoids) such as B-carotene; and polyterpenes (C₄₀₊ isoprenoid compounds derived from more than 8 isoprene units) such as polyisoprene. In some embodiments, the isoprenoid compound or compounds may be any combination of two or more isoprenoid compounds.

[0043] Disclosed herein is a method for producing at least one bio-organic compound is provided which comprises:

- a. culturing, in an aqueous medium, a plurality of genetically modified bacterial or fungal host cells that produce at least one bio-organic compound having at least ten carbon atoms, wherein the aqueous medium forms a first phase, and wherein cells in the plurality-convert a carbohydrate carbon source present in the aqueous medium into the at least one bio-organic compound having at least ten carbon atoms;
- b. forming a liquid organic second phase, comprising the at least one bio-organic compound in an amount of 40-99%, in contact with the first phase;
- c. separating at least a portion of the second phase from the first phase; and
- d. isolating the at least one bio-organic compound from the second phase; wherein the at least one bio-organic compound is an isoprenoid compound; and

wherein the organic second phase is spontaneously formed.

[0044] The isoprenoid production system may include one or more additional processing components including: 1) one or more separation systems for separating the at least one bio-organic compound from the aqueous media and the organic second phase; 2) one or more reactors for biologically or chemically altering the at least one bio-organic compound such as by addition, substitution, hydrogenation, alkylation, hydroxylation, condensation, halogenation or any other suitable reaction; 2) one or more blending vessels or systems for blending the at least one bio-organic compound with one or more additional components; 3) and one or more additional purification or separation systems for further purifying the bio-organic composition or the at least one bio-organic compound.

[0045] The second phase comprises the at least one bio-organic compound. The bio-organic compound forms at least 90% of the second phase. In certain embodiments, the second phase consists essentially of the bio-organic compound.

[0046] In some embodiments, the plurality of host cells includes more than one type of host cell, such as more than one species or strain of host cells, for example 2-5 species or strains of host cells, for example 2, 3, 4 or 5 species or strains of host cells. In some embodiments the plurality of host cells may produce more than one bio-organic compound, such as 2-5 bio-organic compounds, for example 2, 3, 4, or 5 bio-organic compounds.

[0047] The bio-organic compound or compounds may be isolated from the first phase and/or second phase using any suitable separation method. In some embodiments, the bio-organic compound is isolated from the second phase such that it is substantially pure.

[0048] The organic second phase occurs spontaneously as a result of chemical and molecular interactions such as differences in solubility, or hydrophobicity, density, concentration or any other spontaneous phase separation mechanism.

[0049] Once phase separation occurs, the separate phases can be individually drawn from the separation vessel. Any amount of the second phase can be separated from the first phase, e.g. all, a portion, 1% to 100% such as 5% to 95%, 10% to 90%, 20% to 80%, 25% to 75%, 35% to 65%, or 40% to 50% of the second phase may be separated from the first phase. If the organic second phase is less dense than the aqueous first phase, then one or more taps can be provided or placed on the separation vessel near the interface between the two phases (preferably within the organic second phase) to decant the organic second phase before removing the denser aqueous phase. Alternatively, the aqueous first phase can be removed from the separation vessel using an outlet near the bottom of the separation vessel until the organic second phase appears. At which point, the organic second phase can be transferred into a separate location for further processing or storage. Both of the aqueous first and organic second phases can flow out of the separation vessel under the force of gravity, gas pressure or through the use of a pump or pumps or a combination thereof.

[0050] If the organic second phase is denser than the aqueous first phase, then one or more taps can be provided or placed on the separation vessel near the interface between the two phases (preferably within the organic second phase) to decant the aqueous first phase before removing the denser organic second phase. Alternatively the organic second phase may be removed from the separation vessel using an outlet near the bottom of the separation vessel.

[0051] For a continuous process in which the aqueous first phase is denser than the organic second phase, a separation vessel with one or more taps can contain a specified volume of the fermentation medium and host cells, and the continually-produced organic second phase may be decanted through the taps to storage or further processing. If the organic second phase is denser than the aqueous first phase, the organic second phase can be removed continuously from the bottom of the separation vessel at a rate that prevents complete depletion of the organic second phase from the separation vessel to avoid drawing from the aqueous first phase.

[0052] In some embodiments, the bio-organic compound may be isolated from the organic second phase using adsorption, a process in which molecules move from a bulk liquid onto the surface of adsorbents. Illustrative examples of adsorbents include activated carbon; aluminas; aluminosilicates such as zeolites; clays such as fuller's earth; molecular sieves; organic polymers such as polystyrene and resins; and silicas such silica gel. Depending on the adsorbent used, the adsorbent may be used to capture the desired bio-organic product or unwanted byproducts. Isolation by adsorption may be performed using a batch, continuous or semi-continuous process.

[0053] In other embodiments, the bio-organic compound may be isolated from the organic second phase using distillation, a method of separating substances based on differences in their volatilities. In batch distillation; an entire batch of liquid is initially charged to a vessel and then heated or reduced in pressure within the vessel. Vapor is thereby continuously generated and may be condensed to form a liquid distillate which is collected. In continuous equilibrium distillation, a continuously flowing liquid feed is heated or reduced in pressure so as to cause partial vaporization of the mixture and separate recovery of liquid and vapor components. The liquid and vapor disengage while flowing through a distillation column, and the products emerge as vapor and liquid streams. When the vapor and liquid approach phase equilibrium, this is called a flashing process. If desired, the vapor product can be condensed to form a liquid distillate.

[0054] In other embodiments, the bio-organic compound or compounds are isolated from the organic second phase using gas-liquid extraction. This process is also known as stripping and is the transfer of a component dissolved in a liquid stream into a vapor stream in a more concentrated form. Temperature and pressure can be optimized for the transfer of the desired bio-organic compound. Illustrative examples of vapor streams include air and steam. Typically, the liquid stream flows down a column while the vapor stream is bubbled up (flowing countercurrently to the liquid stream).

[0055] In other embodiments, the bio-organic compound is isolated from the organic second phase using liquid-liquid extraction. Also known as solvent extraction, liquid-liquid extraction is the transfer of a substance from one liquid phase into another immiscible liquid phase.

[0056] In a batch liquid-liquid extraction system, the feed liquid (the organic second phase) is mixed with a second immiscible liquid phase in a suitable vessel. The mixture is then permitted to settle into layers and separate into extract and raffinate and the lighter layer can be decanted from the vessel. The desired bio-organic compound or compounds can be in the extract or raffinate depending on the product and solvent used.

[0057] In a continuous liquid-liquid extraction system, differences in density, vapor pressure at a given temperature, or boiling points are used to separate the desired bio-organic product from the feed liquid (the organic phase). Such systems can use mixer/settler tanks, towers or columns, centrifuges and combinations thereof to effect separation.

[0058] In other embodiments, the bio-organic compound is isolated from the organic second and /or the aqueous first phase using ultrafiltration, a pressure-driven membrane process used to separate solution components on the basis of molecular size and shape. Under an applied pressure difference across an ultrafiltration membrane, solvent and small solute species pass through the membrane and are collected as permeate while larger solute species are retained by the membrane and recovered as a concentrated retentate. Ultrafiltration involves solutes whose molecular dimensions are ten or more times larger than those of the solvent and are usually below % micron in size. The solutes or the materials to be separated usually have molecular weights greater than 500 amu, such as macromolecules, colloidal dispersions, and emulsions. A non-limiting example of an ultrafiltration system is a tangential flow ultrafiltration system.

[0059] In some embodiments, the host cells are capable of producing from about 10 to about 50 grams, more than about 15 grams, more than about 20 grams, more than about 25 grams or more than about 30 grams of bio-organic compound per liter of fermentation medium.

[0060] In some embodiments, the host cells are capable of producing from about 50 to about 1500 milligrams, such as more than about 100 milligrams, more than about 150 milligrams, more than about 200 milligrams, more than about 250 milligrams, more than about 500 milligrams, more than about 750 milligrams or more than about 1000 milligrams of bio-organic compound per gram of dry cell weight.

Fuel Composition Production System

[0061] Also described herein is a fuel composition production system comprising:

- a. at least one vessel having a capacity of at least 100 liters;
- b. an aqueous medium, within the vessel, comprising a first phase;
- c. a plurality of host cells, within the aqueous medium, capable of making, producing or synthesizing at least one bio-organic compound; and,
- d. a liquid organic second phase comprising the at least one bio-organic compound in contact with the first phase.

The fuel composition production system may include one or more additional processing components including: 1) one or more separation systems for separating the at least one bio-organic compound from the aqueous media and the organic second phase; 2) one or more reactors for biologically or chemically altering the at least one bio-organic compound such as by addition, substitution, hydrogenation, alkylation, hydroxylation, condensation, halogenation or any other suitable reaction; 2) one or more blending vessels or systems for blending the at least one bio-organic compound with one or more additional fuel components such as a petroleum-based fuel, a fuel additive or a combination thereof; and, 3) one or more additional purification or separation systems for further purifying the fuel composition or the at least one bio-organic compound.

[0062] In some embodiments, the fuel additive is selected from the group consisting of oxygenates, antioxidants, environmental protectants, thermal stability improvers, cetane improvers, stabilizers, cold flow improvers, combustion improvers, anti-foams, anti-haze additives, corrosion inhibitors, lubricity improvers, icing inhibitors, injector cleanliness additives, smoke suppressants, drag reducing additives, metal deactivators, dispersants, detergents, deemulsifiers, dyes, markers, static dissipaters, biocides and combinations thereof.

[0063] Disclosed herein is a fuel composition production system comprising:

- a. one or more fermentation systems comprising:

- i) at least one vessel having a capacity of at least 100 liters;
- ii) an aqueous medium, within the at least one vessel, forming a first phase;
- iii) a plurality of genetically modified bacterial or fungal host cells, within the aqueous medium, capable of converting a carbohydrate carbon source present in the aqueous medium into at least one bio-organic compound having at least ten carbon atoms; and,
- iv) a liquid organic second phase, comprising the at least one bio-organic compound in an amount of 40-99%, in contact with the first phase;
- v) wherein the at least one bio-organic compound is an isoprenoid compound; and

wherein the organic second phase is spontaneously formed;

- b. one or more first phase separation systems whereby the first phase and the second organic phase or one or more components of the second organic phase are separated;
- c. optionally one or more second phase separation systems whereby the at least one bio-organic compound is separated from the second organic phase;
- d. optionally one or more reactors or vessels wherein the at least one bio-organic compound is chemically or biologically modified;
- e. optionally one or more purification systems whereby the bio-organic compound or the modified bio-organic compound is purified or further purified;
- f. optionally one or more blending vessels or systems for blending the at least one bio-organic compound with one or more additional fuel components; and
- g. optionally one or more further purification systems whereby the blend of the at least one bio-organic compound and the one or more additional fuel components is purified or further purified.

[0064] In some embodiments, the one or more first phase separation systems comprises one or more systems, vessels or other phase separation components detailed herein configured specifically to separate the first phase from the second organic phase. In some embodiments the one or more second phase separation systems includes one or more systems, vessels or phase separation components detailed herein configured specifically to separate the bio-organic compound or compounds from the second organic phase.

[0065] In some embodiments, the one or more reactors wherein the at least one bio-organic compound is chemically or biologically modified comprises the same or different vessel or vessels used for the fermentation or the separation systems. Alternatively, the one or more reactors may comprises one or more different vessels, which may include additional hardware, sensors, ports, probes, and/or control systems suitable for the specific reaction or reactions or other modifications to the bio-organic compound or compounds that are performed therein. The reactors may be batch, fed batch or continuous reactors.

[0066] In some embodiments, the bio-organic compounds or modified bio-organic compounds or the fuel compositions

may be purified or further purified using one or more purification systems. The purification systems may comprise any suitable purification system including any system that may remove unwanted compounds from the bio-organic compound or compounds or that may separate the unwanted compounds from the bio-organic compounds. In some embodiments, the purification system may comprise one or more systems, vessels or phase separation components detailed herein that may be specifically configured to achieve the desired purity of the bio-organic compound or compounds. In some embodiments, the purification may be accomplished using one or more separation systems in series to achieve the desired purity. In some embodiments, the separation systems may be configured differently from each other in order to achieve the purity in stepwise fashion.

[0067] In some embodiments, the purification will be performed to achieve specifications or requirements of federal, state or local laws, rules or regulations for the bio-organic compounds or for fuel compositions. In some embodiments, the purification can improve the functionality of the bio-organic compounds or fuel compositions beyond the requirements of federal or state laws, rules or regulations. In some embodiments, the federal state or local laws, rules or regulations may pertain to environmental emissions, fuel performance, tax incentives, and other economic incentives. In some embodiments, the purification may reduce the environmental impact of, carbon footprint of, fuel efficiency obtained from; reliability obtained from, energy available from, or long term economic cost of the bio-organic compounds or fuel compositions.

[0068] In some embodiments the fuel composition system includes one or more blending vessels or systems for blending the at least one bio-organic compound with one or more additional fuel components. The blending vessel or blending system may be any suitable vessel or system. The blending vessel may include any or all of the inlets, outlets, ports, sensors, probes, agitators and other hardware identified for the bio-organic compound production vessel. The blending vessel may blend one or more fuel components with the bio-organic compound or compounds. For example, 2-5 fuel components, such as 3 or 4 fuel components. The blending system may be batch, continuous or fed batch.

[0069] Disclosed herein is a method of making a fuel composition comprising:

- a. culturing, in an aqueous medium, a plurality of genetically modified bacterial or fungal host cells that convert a carbohydrate carbon source in the aqueous medium into at least one bio-organic compound having at least ten carbon atoms, wherein the aqueous medium forms a first phase; wherein the at least one bio-organic compound is an isoprenoid compound;
- b. forming a liquid organic second phase, comprising 40-99% of the at least one bio-organic compound, in contact with the first phase; wherein the organic second phase is spontaneously formed;
- c. separating at least a portion of the second phase from the first phase;
- d. isolating the at least one bio-organic compound from the second phase;
- e. optionally chemically or-biologically modifying the at least one bio-organic compound;
- f. optionally purifying the bio-organic compound or the modified bio-organic compound;
- g. optionally blending the at least one bio-organic compound with one or more additional fuel components; and
- h. optionally purifying the blend of the one or more bio-organic compounds and the one or more additional fuel components.

[0070] In some embodiments, the fuel composition comprises a biofuel composition. In some embodiments, the biofuel further comprises at least one bio-organic compound and a petroleum-based fuel, a fuel additive or a combination thereof. In further embodiments, the petroleum-based fuel is a gasoline, jet fuel, kerosene, diesel fuel or a combination thereof.

[0071] In some embodiments, the bio-organic compound production system or the fuel composition production system may be built or created by retrofitting an ethanol production facility.

[0072] In some embodiments, the fuel composition production systems may comprise one or more automated control systems. The automated control systems may be the same or different from the control systems for the bio-organic production system and may comprise various sensors, probes and other equipment for measuring and controlling the various process parameters associated with each system within the fuel composition system and each step or the fuel composition production methods. The automated system or systems may additionally be controlled and monitored by a central control system, which may be a local or plant wide control system and may control production of just one bio-organic compound production process or multiple bio-organic compound production processes. The automated system or systems and central control system may comprise any suitable software, firmware and/or hardware, which may be proprietary or off the shelf or a combination thereof and may communicate using any suitable communication system. Non-limiting examples of such communication systems include hardwired systems that may be digital or analog, and may include direct connection or be in the form of a network such as a LAN or a WAN or ethernet. In addition, in some embodiments the communication system may be wireless and may be proprietary, BLUETOOTH, ultra wide band, 802.11 a,b,g or n or ZigBee, including TDMA, FDMA, OFDM, and CDMA and may operate in any suitable frequency band such as 2.4 GHz or 5.8 GHz.

Host Cells

[0073] Genetically modified bacterial or fungal host cells can be used in the practice of the present invention. In some embodiments, 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 bio-organic compound, or effect an increased yield of the desired bio-organic compound.

[0074] Examples of bacterial cells 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*, *Scenedesmus*, *Serratia*, *Shigella*, *Staphylococcus*, *Streptomyces*, *Synnecoccus*, and *Zymomonas*.

[0075] Illustrative examples of bacterial species include but are not limited to: *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *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.

[0076] In general, if a bacterial host cell is used, a non-pathogenic strain is preferred. Illustrative examples of species with non-pathogenic strains include but are not limited to: *Bacillus subtilis*, *Escherichia coli*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Pseudomonas putida*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, and the like.

[0077] Examples of fungal cells include, but are not limited to those belonging to the genera: *Aspergillus*, *Candida*, *Chrysosporium*, *Cryptococcus*, *Fusarium*, *Kluyveromyces*, *Neotryphodium*, *Neurospora*, *Penicillium*, *Pichia*, *Saccharomyces*, *Trichoderma* and *Xanthophyllomyces* (formerly *Phaffia*)

[0078] - Illustrative examples of fungal species 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*, *Saccharomyces bayanus*, *Saccharomyces 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*).

[0079] In general, if a fungal cell is used, a non-pathogenic strain is preferred. Illustrative examples of species with non-pathogenic strains include but are not limited to: *Fusarium graminearum*, *Fusarium venenatum*, *Pichia pastoris*, *Saccharomyces-boulardi*, and *Saccharomyces cerevisiae*.

[0080] In some embodiments, the host cells of the present invention have been designated by the Food and Drug Administration as GRAS or Generally Regarded As Safe. Illustrative examples of such strains include: *Bacillus subtilis*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, and *Saccharomyces cerevisiae*.

Engineering Pathways to make Bio-Organic Compounds

[0081] An illustrative example of a class of bio-organic compounds is isoprenoids. 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. In addition to their usefulness in making fuel compositions, a vast number of natural and synthetic isoprenoids are useful as pharmaceuticals, cosmetics, perfumes, pigments and colorants, fungicides, antiseptics, nutraceuticals, and fine chemical intermediates.

[0082] Isoprenoid compounds are made in nature through two different metabolic pathways which converge at 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. The methods described herein for engineering the MEV and DXP pathways to make the desired isoprenoid compound can be readily adapted to similarly engineer other pathways to make other bio-organic compounds.

MEV Pathway

[0083] . A schematic representation of the MEV pathway is described in Figure 3. In general, the pathway comprises

six steps.

[0084] 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*).

[0085] 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*).

[0086] 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*), (AX1282-13, providing the sequence encoding a truncated HMGR; *Saccharomyces cerevisiae*), and (NC_001145: complement (115734..118898; *Saccharomyces cerevisiae*).

[0087] 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*).

[0088] 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*).

[0089] 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*).

[0090] 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

[0091] A schematic representation of the DXP pathway is described in Figure 4. 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* Temecula), and (NC_003076, locus tag-AT5G11380; *Arabidopsis thaliana*).

[0092] 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).

[0093] 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).

[0094] 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).

[0095] 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).

[0096] 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).

[0097] 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).

[0098] 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 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*.

[0099] 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 naturally occurring DXP pathway enzymes.

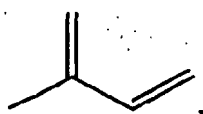
[0100] In other embodiments, the host cell produces IPP via the DXP pathway, either exclusively or in combination with the MEV pathway. In other embodiments, a host's MEV pathway is functionally disabled so that the host cell produces IPP exclusively through a heterologously introduced DXP pathway. The MEV pathway can be functionally disabled by disabling gene expression or inactivating the function of one or more of the naturally occurring MEV pathway enzymes.

C₅ Compounds

[0101] Exemplary C₅ bio-organic compounds are hemiterpenes which are generally derived from IPP or DMAPP. An illustrative example of a hemiterpene is isoprene.

Isoprene

[0102] 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: (AB 198190; *Populus alba*) and (AJ294819; *Populus alba* x *Populus tremula*).

C₁₀ Compounds

[0103] Exemplary C₁₀ bio-organic compounds are monoterpenes which are generally derived from geranyl pyrophosphate (GPP) which in turn is made by the condensation of IPP with DMAPP. An enzyme known to catalyze this step is, for example, geranyl pyrophosphate synthase.

[0104] Figure 5 shows schematically how IPP and DMAPP can produce GPP, which can be further processed to a monoterpene.

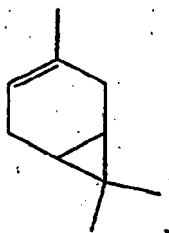
[0105] Illustrative examples of nucleotide sequences for geranyl pyrophosphate synthase include but are not limited to: (AF513111; *Abies grandis*), (AF513112; *Abies grandis*), (AF513113; *Abies grandis*), (AY534686; *Antirrhinum majus*),

(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*), (AF 182828; *Mentha x piperita*), (AF182827; *Mentha x piperita*), (MPI24945; *Mentha x piperita*), (PZE431697, Locus CAD24425; *Paracoccus zeaxanthinifaciens*), (AY866498; *Picrorhiza kurrooa*), (AY351862; *Vitis vinifera*), and (AF203881, Locus AAF12843; *Zymomonas mobilis*).

[0106] GPP is then subsequently converted to a variety of C₁₀ compounds. Illustrative examples of C₁₀ compounds include but are not limited to:

Carene

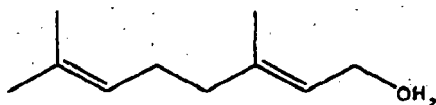
[0107] Carene, whose structure is



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*).

Geraniol

[0108] 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*).

Linalool

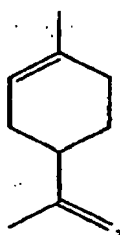
[0109] Linalool, whose structure is



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).

Limonene

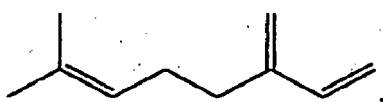
[0110] Limonene, whose structure is



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*).

Myrcene

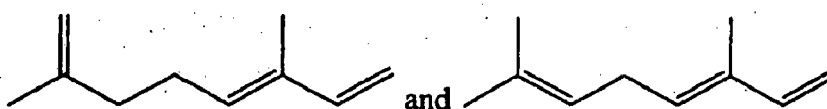
[0111] Myrcene, whose structure is



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*).

Ocimene

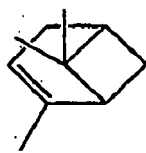
[0112] α - and β -Ocimene, whose structures are



respectively, 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).

 α -Pinene

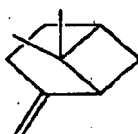
[0113] α -Pinene, whose structure is



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*).

β -Pinene

[0114] 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*).

Sabinene

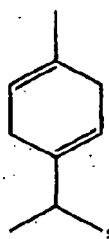
[0115] Sabinene, whose structure is



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*.

γ -Terpinene

[0116] γ -Terpinene, whose structure is

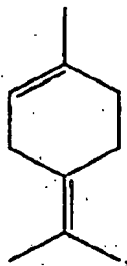


is a constituent of the essential oil from citrus fruits. Biochemically, γ -terpinene is made from GPP by a γ -terpinene

synthase. Illustrative examples of suitable nucleotide sequences include: (AF514286, REGION: 30..1832 from *Citrus limon*) and (AB110640, REGION 1..1803 from *Citrus unshiu*).

Terpinolene

[0117] Terpinolene, whose structure is



is found in black currant, cypress, guava, lychee, papaya, pine, and tea. Terpinolene is made from GPP by terpinolene synthase. Illustrative examples of a suitable nucleotide sequence include but is not limited to: (AY693650 from *Oscimum basilicum*) and (AY906866, REGION: 10..1887 from *Pseudotsuga menziesii*).

C₁₅ Compounds

[0118] Exemplary C₁₅ bio-organic compounds are sesquiterpenes which are generally derive from farnesyl pyrophosphate (FPP) which in turn 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.

[0119] Figure 5 also shows schematically how IPP and DMAPP can be combined to produce FPP, which can be further processed to a sesquiterpene.

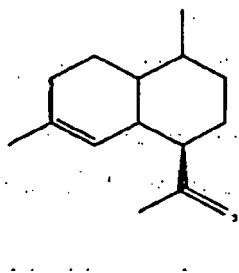
[0120] Illustrative example of nucleotide sequences for farnesyl pyrophosphate synthase include but are not limited to: (ATU80605; *Arabidopsis thaliana*), (ATHFPS2R; *Arabidopsis thaliana*), (AAU36376; *Artemisia annua*), (AF461050; *Bos taurus*), (D00694; *Escherichia coli* K-12), (AE009951, Locus AAL95523; *Fusobacterium nucleatum subsp. nucleatum* ATCC 25586), (GFFPPSGEN; *Gibberella fujikuroi*), (CP000009, Locus AAW60034; *Gluconobacter oxydans* 621H), (AF019892; *Helianthus annuus*), (HUMFAPS; *Homo sapiens*), (KLPFPSQCR; *Kluyromyces 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* MGAS 10750), and (MZEFPF; *Zea mays*).

[0121] 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* Temecula).

[0122] FPP is then subsequently converted to a variety of C₁₅ compounds. Illustrative examples of C₁₅ compounds include but are not limited to:

Amorphadiene

[0123] Amorphadiene, whose structure is

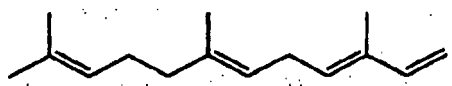


is a precursor to artemisinin which is made by *Artemisia anna*. Amorphadiene is made from FPP by amorphadiene synthase. An illustrative example of a suitable nucleotide sequence is SEQ ID NO. 37 of U.S. Patent Publication No. 2004/0005678.

[0124] Figure 5 shows schematically how IPP and DMAPP can be combined to produce FPP, which can then be further processed to produce amorphadiene.

α -Farnesene

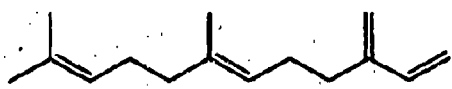
[0125] α -Farnesene, whose structure is



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 AFS 1). Pechouus et al., Planta 219(1):84-94 (2004).

β -Farnesene

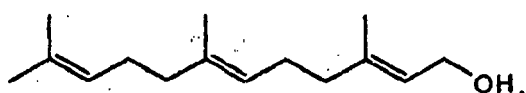
[0126] β -Farnesene, whose structure is



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 A024615 from *Mentha x piperita* (peppermint; gene Tspa11), and AY835398 from *Artemisia annua*. Picaud et al., Phytochemistry 66(9): 961-967 (2005).

Farnesol

[0127] Farnesol, whose structure is

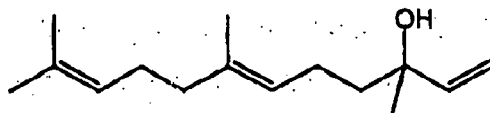


is found in various biological sources including insects and essential oils such as from citronella, neroli, cyclamen, lemon grass, tuberose, and rose. Farnesol is made from FPP by a hydroxylase such as farnesol synthase. Illustrative examples of suitable nucleotide sequences include but are not limited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

nology 128:149-158 (2006).

Nerolidol

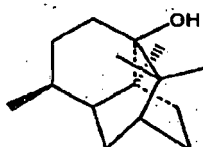
[0128] Nerolidol, whose structure is



is also known as peruvial, 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).

Patchoulol

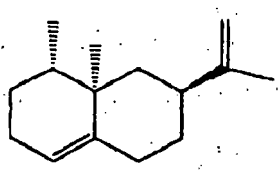
[0129] Patchoulol, whose structure is



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

Valenecene

[0130] 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*.

C₂₀ Compounds

[0131] Exemplary C₂₀ bio-organic compounds are diterpenes which are generally derived from geranylgeraniol pyrophosphate (GGPP) which in turn 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.

[0132] Figure 5 also 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.

[0133] 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), (CRG-

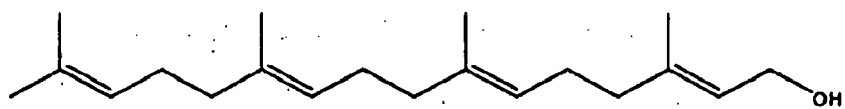
GPPS; *Cathararithus roseus*), (NZ_AABF02000074, Locus ZP_00144509; *Fusobacterium nucleatum subsp. vincentii*, ATCC 49256), (GFGGPPSGN; *Gibberella fujikuroi*), (AY371321; *Ginkgo biloba*), (AB055496; *Hevea brasiliensis*), (AB017971; *Homo sapiens*), (MCI276129; *Mucor circinelloides f. lusitanicus*), (AB016044; *Mus musculus*), (AABX01000298, Locus NCU01427; *Neurospora crassa*), (NCU20940; *Neurospora crassa*), (NZ_A.AKL01000008, Locus ZP_00943566; *Ralstonia solanacearum* UW551), (AB118238; *Rattus norvegicus*), (SCU31632; *Saccharomyces cerevisiae*), (AB016095; *Synechococcus elongates*), (SAGGPS; *Sinapis alba*), (SSOGDS; *Sufolobus acidocaldarius*), (NC_007759; Locus YP_461832; *Syntrophus aciditrophicus* SB), and (NC_006840, Locus YP_204095; *vibrio fischeri* ES 14).

[0134] 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*), (ER-WCRTE; *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).

[0135] GGPP is then subsequently converted to a variety of C₂₀ isoprenoids. Illustrative examples of C₂₀ compounds include but are not limited to:

Geranylgeraniol

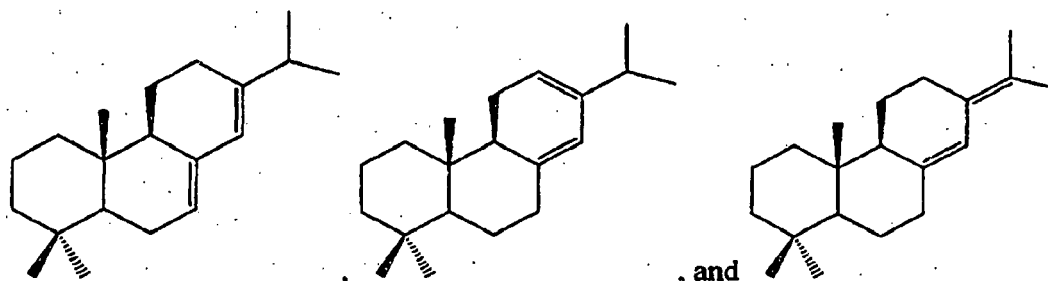
[0136] Geranylgeraniol, whose structure is



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.

Abietadiene

[0137] 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

[0138] C₂₀₊ bio-organic 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

[0139] Although for illustrative purposes, the invention has been described with reference to engineering the MEV and/or DXP pathways, these methods can be adapted to similarly engineer suitable pathways to make non-isoprenoid

bio-organic compounds. These pathways are typically engineered, using recombinant DNA technology by expression of suitable heterologous sequences encoding one or more enzymes.

[0140] The subject nucleotide acids can be expressed by a single or multiple vectors. The nucleic acids can be arranged in a single operon, or in separate operons that are placed in one or multiple vectors. Where desired, two expression vectors can be employed, each of which contains one or more heterologous sequences operably linked in a single operon. While the choice of single or multiple vectors and the use of single or multiple operons may depend on the size of the heterologous sequences and the capacity of the vectors, it will largely depend on the overall yield of a given bio-organic compound that the vector is able to provide when expressed in a selected host cell. In some instances, two-operon expression system provides a higher yield of the bio-organic compound. 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.

[0141] In certain host cells, the subject nucleic acids may be controlled by one or more operons. In some instances, a two or three operon system provides a higher yield of a bio-organic compound over a single operon system.

[0142] 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.* 275(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 for use in 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.

[0143] 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.

[0144] 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.).

[0145] 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 RNA 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 including 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.

[0146] 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 desired enzyme. 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.

[0147] Suitable low copy expression vectors for *Escherichia coli* include, but are not limited to, pAGYC184, 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 (Erhardt 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).

[0148] Regulatory elements include, for example, promoters and operators can also be engineered to increase the metabolic flux of the engineered pathways by increasing the expression of one or more genes that play a significant role in determining the overall yield of the bio-organic compound 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.

[0149] 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.

[0150] 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. USA. 89(21):10079-83), a *nirB* promoter (Harbome 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.

[0151] In some embodiments, the total activity of a heterologous enzyme that plays a larger role in the overall yield of a bio-organic compound 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 one or more engineered 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.

[0152] 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.

[0153] 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 exam-

ple, 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.

[0154] . 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 comprises a promoter that is stronger than a native E. Coli Lac promoter.

[0155] 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.)

[0156] 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).

[0157] 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.

[0158] 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.

[0159] 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 example 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.

[0160] For production of a bio-organic product 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.

[0161] 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.

[0162] 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.

[0163] The introduction of various pathway sequences into a host cell for use in the invention 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.

[0164] The yield of a bio-organic compound via one or more metabolic pathways disclosed herein can be augmented by inhibiting reactions that divert intermediates from productive steps towards formation of the bio-organic 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 the bio-organic compound.

[0165] 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.

[0166] 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 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.

EXAMPLES

[0167] 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.

[0168] 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.

Example 1

[0169] This example describes methods for making expression plasmids that encode enzymes of the MEV pathway from *Saccharomyces cerevisiae* organized in operons.

[0170] 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_C00913 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}Nsil-XmaI* fragment of pBAD33 was replaced with the *Nsil-XmaI* fragment of pBBR1MGS, yielding expression plasmid pMevT (see U.S. Patent Number 7,192,751).

[0171] 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 enzymes 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.

[0172] 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.

[0173] 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 *MVD1* 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): 175-175 (1995)), yielding expression plasmid pMevB-Cm.

[0174] 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 AF 119715) 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).

[0175] Expression plasmid pMBIS was generated by inserting the *ispA* gene into pMBI. The *ispA* gene encodes a farnesyl pyrophosphate synthase that catalyzes the conversion of IPP and DMAPP to 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).

[0176] 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 Y 17376, 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 6 for a plasmid map).

[0177] 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 plasmids 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

[0178] This example describes methods for making expression vectors encoding enzymes of the MEV pathway from *Staphylococcus aureus* organized in operons.

[0179] 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. *ERG13* is also known as HMGS or HMG-CoA synthase.

[0180] 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). 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. (Bio Techniques 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.

[0181] 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.

[0182] 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.

[0183] 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 pAM45 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.

[0184] 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

[0185] This example describes methods for making expression vectors that encode enzymes of the MEV pathway from *Enterococcus faecalis* organized in operons.

[0186] 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.

[0187] 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 BgIII (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.

[0188] Expression plasmid pAM22 was generated by inserting the coding sequence of the *mvaE* gene of expression plasmids 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 *BamHI* *HindIII* restriction enzyme site of pZE21-P_{L-lacO1}, yielding expression plasmid pAM22.

[0189] 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 *BgIII* and *BamHI* 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 *BamHI* site of expression plasmid pAM22, yielding expression plasmid pAM33.

[0190] 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 *EcoRI* restriction enzyme, digesting the resulting linear DNA fragment using *MluI* 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 *MluI* and *EcoRI* 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

[0191] This example describes methods for making expression plasmids that encode enzymes of the DXP pathway from *Escherichia coli* organized in operons.

[0192] 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 (*XhoI* and *KpnI* for the PCR product comprising the *dxs* gene; *KpnI* and *Apal* for the PCR product comprising the *dxr* gene; *Apal* and *NdeI* for the PCR product comprising the *ispD* gene; *NdeI* and *MluI* 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 μ l 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 67-1 E-C (SEQ ID NO: 19) and 67-1 H-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 *Apal* restriction enzymes, and the PCR product comprising the *ispD-ispE* gene cassette was digested to completion using *Apal* 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 7 for a plasmid map).

[0193] 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 *Apal* for the PCR product comprising the *ispF* gene; *KpnI* and *Apal* 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 μ l 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 oligonu-

cleotide-generated lacUV5 promoter. Vector pAM369 was digested to completion using *Bam*HI and *Nco*I restriction enzymes, and the 2 isolated PCR products containing the "bottom" DXP pathway operon were ligated into the *Bam*HI *Nco*I restriction enzyme site of the pAM369 vector, yielding expression plasmid pAM409.

[0194] Expression plasmid pAM424, a derivative of expression plasmid pAM409 containing the broad-host range RK2 origin of replication, was generated by transferring the *lac*UV5 promoter and the *isp*FGH-*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 *Aat*II and *Xho*I 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 *Sac*I and *Hind*III, 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 *Sac*I *Hind*III 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 *Bgl*II and *Hind*III 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 *Pci*I and *Xho*I restriction enzyme sites, yielding vector pAM257. Expression plasmid pAM409 was digested to completion using *Xho*I and *Pci*I 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 *Xho*I and *Pci*I, and the isolated DNA fragment containing the *lac*UV5 promoter and *isp*FGH-*idi-ispA* operon was ligated into the *Xho*I *Pci*I restriction enzyme site of the pAM257 vector, yielding expression plasmid pAM424 (see Figure 8 for a plasmid map).

Example 5

[0195] This example describes methods for making expression plasmids that encode enzymes that convert FPP or GPP.

[0196] Expression plasmid pTrc99A-ADS was generated by inserting a nucleotide sequence encoding an amorpho-4,11-diene synthase ("ADS") into vector pTrc99A. The amorpho-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 amorpho-4,11-diene synthase was optimized for expression in *Escherichia coli*, and so that the nucleotide sequence was flanked by a 5' *Nco*I and a 3' *Xma*I restriction enzyme site (see U.S. Patent Number 7,192,751). The nucleotide sequence was digested to completion using *Nco*I and *Xma*I 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 *Nco*I *Xma*I restriction enzyme site of the pTrc99A vector (Amman et al. (1985) *Gene* 40:183-190), yielding expression plasmid pTrc99A-ADS (see Figure 9 for a plasmid map).

[0197] 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 *Fsp*I restriction enzyme site of expression plasmid pTrc99A-ADS, yielding expression plasmid pAM113.

[0198] 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* 69:301-3115). 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.

[0199] 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 *Nco*I 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 *Nco*I restriction enzyme, and digested to completion using *Sac*I 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 9 for a plasmid map).

[0200] 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 9 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 9 for plasmid maps).

[0201] 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.

[0202] Expression plasmids pTrc98A-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 *Xma*I *Xba*I restriction enzyme site (for the β -pinene, sabinine, and β -phellandrene synthase inserts), the *Nco*I *Xma*I restriction enzyme site (for the linalool and carene synthase inserts), or the *Xba*I *Pst*I 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 9 for plasmid maps).

Example 6

[0203] This example describes the generation of *Escherichia coli* host strains useful in the invention.

[0204] 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 | |
| B210 | DP | pTrc99A-ADS | |
| 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 |
| B61 | DH1 | pAM25 pBBR1MCS-3 pTrc99A | 5 ug/mL tetracycline |
| B62 | | pAM34 pBBR1MCS-3 pTrc99A | |
| B003 | DH10B | pTrc99A-ADS | 100 μ g/ml carbenicillin |
| B617 | | pAM408 pTrc99A-ADS | 100 ug/mL carbenicillin 50 ug/mL kanamycin |
| B618 | | pAM424 pTrc99A-ADS | 100 ug/mL carbenicillin 35 μ g/ml chloramphenicol |
| B619 | | pAM408 pAM424 pTrc99A-ADS | 100 μ g/ml carbenicillin 50 μ g/ml kanamycin 35 μ g/ml chloramphenicol |

(continued)

| Host Strain | E.coli Parent Strain | Expression Plasmids | Antibiotic Selection |
|---|----------------------|--|---|
| B650 | DH10B | pAM373 | 100 µg/ml carbenicillin |
| B651 | | pAM408 | 100 µg/ml carbenicillin |
| B652 | | pAM373 | 50 µg/ml kanamycin |
| B653 | | pAM424 | 100 µg/ml carbenicillin |
| | | pAM373 | 35 µg/ml chloramphenicol |
| B286 | DH1 | pAM408 pAM424 pAM373 | 100 µg/ml carbenicillin 50 µg/ml kanamycin 35 µg/ml chloramphenicol |
| B287 | | pAM97-MevB pC9 | 100 ug/mL carbenicillin |
| B288 | | pAM97-MevB pnudF-H | 34 ug/mL chloramphenicol. |
| B291 | | pAM97-MevB pyhfR | |
| B592 | DH1 | pAM97-MBI pyhfR | |
| B552 | | pMevT pMBIS pTrc99A-FSA | 100 ug/mL carbenicillin 34 ug/mL chloramphenicol |
| Example 21 host cell (production of GTS, APS, TS) | | pMevT pMBIS pAM373 | 5 ug/mL tetracycline |
| Example 21 host cell (production of LLS, LMS, BPS, PHS, CS, SS) | | pMevT pMBIS-gpps pTrc99A-GTS or -APS or -TS | |
| | | 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 |

[0205] 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 |

(continued)

| Component | Quantity (per L) |
|--|------------------|
| KH ₂ PO ₄ | 3g |
| 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) 6Mo7O 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

[0206] 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.

[0207] 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.

[0208] 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.

[0209] 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

[0210] 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.

[0211] 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.

[0212] 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 µL per 5 mL bugbuster; Novagen, Madison, WI), and lysozyme (30 µg/mL). Enzyme activity of the *Saccharomyces cerevisiae* tHMGR was assayed in 50 mM Tris HCl (pH 7.5), 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 340 nm. 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 (pH 6.5), 0.4 mM NADPH, 1.0 mM EDTA, and 100 mM KCl.

[0213] Protein assays were done by the method of Bradford ((1976) Anal Biochem. 72:248-254). Specific activities were calculated as Δ nmol NADFH/min/mg protein.

Example 9

[0214] This example describes the calibration of OD₆₀₀ with dry cell weight ("DCW").

[0215] 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 10-12. 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.

Example 10 (comparative example)

[0216] 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.

[0217] 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.

[0218] 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 µL of 1M IPTG to the culture medium. The cultures were overlain with 8 mL 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 µL of the organic overlay layer to a clean glass vial containing 500 µL ethyl acetate spiked with beta- or trans-caryophyllene as an internal standard.

[0219] 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 µL sample was separated on the GC using a DB-XXL 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.

Example 11 (comparative example)

[0220] This example demonstrates increased production of amorpho-4,11-diene by an *Escherichia coli* host strain grown at suboptimal temperature.

[0221] 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.

[0222] 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.

Example 12 (comparative example)

[0223] This example demonstrates increased production of amorpho-4,11-diene by an *Escherichia coli* host strain grown under restricted carbon source conditions.

[0224] 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.

[0225] 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.

[0226] 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 | 5g | 590-650 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 | 2g | - |
| MgSO ₄ 7H ₂ O | 1.2 g | 12 g |
| EDTA | 8.4 mg | 13 g |
| CoCl ₂ 6H ₂ O | 0.25 mg | 0.4 mg |
| MnCl ₂ 4H ₂ O | 1.5 mg | 2.35 mg |
| CuCl ₂ 2H ₂ O | 0.15 mg | 0.25 mg |
| H ₃ BO ₄ | 0.3 mg | 0.5 mg |
| Na ₂ MoO ₄ 2H ₂ O | 0.25 mg | 0.4 mg |
| Zn(CH ₃ COO) ₂ 2H ₂ 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 |

[0227] For fermentation run 050608-1 (excess carbon), the feed was initiated at induction, and feed rates were adjusted manually. 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:

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

$$\mu = 0.12$$

$$S(t_0) = 15g$$

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 |

(continued)

| Run Time (hours) | Glucose Feed Rate (g/hr) |
|------------------|--------------------------|
| 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 |

[0228] 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-Aldrich, 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.

[0229] 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.

[0230] 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 µL broth plus organic overlay with 975 µL 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.

Example 13 (comparative example)

[0231] This example demonstrates increased amorpho-4,11-diene production by an *Escherichia coli* host strain grown under restricted carbon source conditions and at suboptimal temperature.

[0232] 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.

[0233] 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.

[0234] Production of amorpho-4,11-diene in the host cells was induced by adding 1 mL of 1 M IPTG to the culture medium. Amorpho-4,11-diene was captured and extracted according to two different protocols. In one method, 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 and 20 mg/L. In another, amorpho-4,11-diene was captured by adding 200 mL of an organic overlay to the fermentor at the time of induction.

[0235] Amorpho-4,11-diene was extracted from the culture medium by combining 25 µL broth with 975 µL 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.

Example 14 (comparative example)

[0236] 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.

[0237] 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, subcultured 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.

[0238] 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.

[0239] 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:

$$m_x(t) = S_0 \mu e^{\mu(t-t_0)}$$

$$\mu = 0.12 \text{ min}^{-1}$$

$$S_0 = 15 \text{ g}$$

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.

[0240] 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.

[0241] 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.

[0242] 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.

[0243] Samples were analyzed by GC/MS as described in Example 10.

Example 15 (comparative example)

[0244] This example describes the production of amorpho-4,11-diene via the DXP pathway in an *Escherichia coli* host strain.

[0245] 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.

[0246] 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.

[0247] 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%.

Example 16 (comparative example)

[0248] 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.

[0249] 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 inoculate 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 uL 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 uL 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 uL of ethyl acetate was added to 300 uL of each removed sample. The sample was then vortexed for 15 minutes, and 400 uL of the upper ethyl acetate phase was transferred to a clean glass vial for analysis.

[0250] The samples were analyzed on a Hewlett-Packard 6890 gas chromatograph/mass spectrometer (GC/MS). A 1 uL 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 (comparative example)

[0251] This example describes the production of amorpho-4,11-diene by a *Saccharomyces cerevisiae* host strain.

[0252] The generation of host strain EPY224 is described in Ro et al. (Nature 440: 40-943;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.

[0253] 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.

[0254] 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 *GAL1* 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 μ L of the organic solvent layer to a clean glass vial containing 500 μ L ethyl acetate containing a known concentration of beta- or trans-caryophyllene as an internal standard.

[0255] The organic overlay/ethyl acetate samples were analyzed on a Hewlett-Packard 6890 gas chromatograph/mass spectrometer (GC/MS) as described in Example 10.

[0256] After 72 hours of growth, 3 yeast cultures were found to produce 60.68, 54.48, and 59.25 mg/L amorpha-4,11-diene.

Example 18 (comparative example)

[0257] This example describes the production of amorpha-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.

[0258] Yeast strains CEN.PK2-1C (Y002) (MATA; ura3-52; trp1-289; leu2=3,112; his3 Δ 1; MAL2-8C; SUC2) and CEN.PK2-1D (Y003) (MAT α ; ura3-52; trp1-289; leu2-3,112; his3 Δ 1; 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.

[0259] 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 DH5 α .

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 | ERG 13 | URA3 | URA3 |
| pAM492 | URA3 | IDI1 | CUP1 | CUP1 | tHMGR | URA3 | URA3 |
| pAM493 | ADE1 | tHMGR | GAL1 | GAL10 | IDI1 | ADE1 | URA3 |
| pAM494 | ADE1 | tHMGR | CUP1 | CUP1 | IDI1 | 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 |

[0260] *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. 10 μ g of the resulting PCR product was transformed into exponentially growing Y002 and Y003 strains using 40% w/w polyethelene glycol 3350 (Sigma-Aldrich St Louis, MO), 100 mM lithium acetate (Sigma), 10 μ g 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 0.5 μ g/ml Geneticin (Invitrogen Co; Carlsbad, CA) and confirmed by diagnostic PCR. The resultant clones were given the designation Y93 (MAT A) and Y94 (MAT α). 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). 10 μ g 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).

[0261] To generate *S. cerevisiae* strain Y188; 2 μ g'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.

[0262] To generate *S. cerevisiae* strain Y189, 2 μ g'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 Y 177 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.

[0263] Approximately 1 X 10⁷ 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. 2 μ g 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.

[0264] 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).

[0265] 2 μ g of plasmid DNA from pAM426 (SEQ ID NO:53), containing *S. cerevisiae* condon optimized Amorphadeine 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).

[0266] 2 μ g of plasmid DNA from pAM322 (SEQ ID NO: 54), containing *S. cerevisiae* condon optimized Amorphadeine Synthase (ADS) and cytochrome P450 monooxygenase (AMO) expressed from the *S. cerevisiae* GAL1 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 (comparative example)

[0267] This example describes the production of α -farnesene or β -farnesene in *Escherichia coli*. host strains.

[0268] 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.

[0269] 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.

[0270] 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 (3-arnesene 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.

[0271] 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 (comparative example)

[0272] This example describes the production of β -farnesene via the DXP pathway in an *Escherichia coli* host strain.

[0273] 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.

[0274] 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 centrifuge 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 19. *Escherichia coli* host strain B653 produced approximately 7 mg/g DCW β -farnesene.

Example 21 (comparative example)

[0275] This example describes the production of α -farnesene or β -farnesene in a *Saccharomyces cerevisiae* host strain.

[0276] 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.

[0277] 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 cryo-vials in 1 mL frozen aliquots made up of 400 μ L 50% glycerol and 600 μ L liquid culture.

[0278] 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 *Ga11* 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.

[0279] 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 (comparative example)

[0280] This example describes the production of γ -terpinene, α -pinene, and terpinolene in *Escherichia coli* host strains.

[0281] 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.

[0282] 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 μ L of 1 M IPTG. Samples were taken once per day for 96 hours by transferring 200 μ L 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 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. The cell pellets were extracted in the ethyl acetate by vortexing the glass vials for 15 minutes, after which 500 μ L of the ethyl acetate extraction was transferred to a clean glass vial.

[0283] 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 ethyl acetate.

Examples 23

[0284] This example describes the production of linalool, limonene, β -pinene, β -phellandrene, carene, or sabinine in *Escherichia coli* host strains.

[0285] 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.

[0286] 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 (not according to the method of the invention), 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 (according to the method of the invention).

Sequence Listing

[0287]

SEQ ID NO: 1
MevT66 operon

GAATTCAAAGGAGGAAAATAAAATGAAGAACTGTGTGATTGTTTCTGCGGTCCG
CACGGCGATCGGCAGCTTTAACGGCTCTTTAGCGAGCACCTCTGCAATCGATCTG
5 GGTGCGACGGTCATTAAGGCCGCCATTGAACGCGCCAAAATCGACAGCCAGCAC
GTTGATGAGGTGATCATGGGCAATGTGTTACAAGCCGGCCTGGGTCAAAACCCA
GCGCGTCAAGCACTGTATAAAATCTGGTCTGGCCGAGACCGTGTGTGGCTTCACCG
10 TCAATAAGGTTTTCGGCTCTGGCCTGAAGAGCGTGGCCCTGGCAGCACAAAGCGA
TTCAAGCCGGTCAGGCACAAAGCATCGTTGCGGGTGGCATGGAGAACATGTCTC
TGGCGCCGTACTTATTAGATGCCAAAGCCCGCAGCGGTTATCGCCTGGGCGATGG
15 TCAGGTGTACGACGTCTCTTACGCGATGGCTTAATGTGCGCGACCCACGGTTAC
CACATGGGTATTACGGCCGAAAACGTGGCGAAAGAATACGGCATTACGCGCGAG
ATGCAGGATGAATTAGCACTGCACTCTCAGCGCAAAGCAGCAGCCGCGATCGAG
TCTGGTGC GTTTACGGCGGAAATCGTGCCAGTTAACGTGGTCACGCGCAAGAAG
20 ACGTTCGTTTTTCAGCCAGGACGAGTTCCCGAAGGCAAACAGCACCGCGGAGGCC
TTAGGTGCCTTACGCCCAGCCTTTGACAAAGCGGGCACGGTCACCGCCGGTAATG
CGAGCGGCATCAATGATGGTGCAGCGGCACTGGTCATCATGGAAGAGAGCGCCG
25 CATTAGCAGCGGGTCTGACCCCAATTAGCGCGCATTAAATCTTATGCCAGCGGCGG
CGTCCCACCAGCCCTGATGGGCATGGGTCCGGTCCCAGCCACGCAAAAAGCCCT
GCAATTAGCGGGCCTGCAACTGGCCGACATTGATCTGATCGAGGCGAACGAGGC
GTTTGCAGCGCAGTTCCTGGCGGTGGGTAAGAATCTGGGCTTCGACAGCGAGAA
30 AGTCAATGTGAACGGTGGCGCGATTGCGTTAGGCCATCCGATTGGTGCAAGCGG
CGCACGCATCTTAGTGACGTTACTGCACGCCATGCAGGCACGCGACAAGACCTTA
GGCCTGGCGACCTTATGTATTGGTGGCGGTCAAGGTATCGCCATGGTGATCGAAC
35 GCCTGAACTGAAGATCTAGGAGGAAAGCAAAATGAAACTGAGCACCAAGCTGTG
CTGGTGTGGCATCAAGGGTCGCCTGCGCCACAAAAGCAGCAACAGCTGCACAA
CACGAACCTGCAAATGACCGAGCTGAAAAAGCAGAAGACGGCCGAGCAAAAGA
40 CCCGCCCCGAGAACGTTGGCATCAAGGGCATCCAGATTTATATCCCGACGCAGTG
TGTC AACCAATCTGAGCTGGAGAAATTCGATGGCGTCAGCCAGGGTAAGTACAC
CATCGGCCTGGGCCAGACCAACATGAGCTTCGTGAACGACCGTGAGGACATCTA
TTCTATGAGCCTGACGGTGCTGTCTAAGCTGATCAAGAGCTACAACATCGACACG
45

AATAAGATCGGTCGTCTGGAGGTGGGTACGGAGACGCTGATTGACAAGAGCAAA
AGCGTGAAGTCTGTCTTAATGCAGCTGTTTCGGCGAGAACACGGATGTTCGAGGGT
5 ATCGACACCCTGAACGCGTGTTACGGCGGCACCAACGCACTGTTCAATAGCCTGA
ACTGGATTGAGAGCAACGCTGGGATGGCCGCGATGCGATCGTCGTGTGCGGCG
ATATCGCCATCTATGACAAGGGTGCGGCACGTCCGACCGGCGGTGCAGGCACCG
10 TTGCGATGTGGATTGGCCCGGACGCACCAATTGTCTTCGATTCTGTCCGCGCGTC
TTACATGGAGCACGCCTACGACTTTTACAAGCCGGACTTCACGAGCGAATACCCG
TACGTGGACGGCCACTTCTCTCTGACCTGCTATGTGAAGGCGCTGGACCAGGTTT
ATAAGTCTTATAGCAAAAAGGCGATTTCTAAGGGCCTGGTCAGCGACCCGGCAG
15 GCAGCGACGCCCTGAACGTGCTGAAGTATTTGACTACAACGTGTTCCATGTCCC
GACCTGCAAATTAGTGACCAAATCTTATGGCCGCCTGTTATATAATGATTTCCGT
GCCAACCCGCAGCTGTTCCCGGAGGTTGACGCCGAGCTGGCGACGCGTGATTAC
GACGAGAGCCTGACCGACAAGAACATCGAGAAGACCTTCGTCAACGTCGCGAAG
20 CCGTTCCACAAAGAGCGTGTGGCCCAAAGCCTGATCGTCCCGACCAACACGGGC
AACATGTATACCGCGTCTGTCTACGCGGCATTTCGCGAGCCTGCTGAATTACGTCG
GTTCTGACGACCTGCAGGGCAAGCGCGTTGGCCTGTTACGCTACGGTAGCGGCTT
25 AGCGGCCAGCCTGTATAGCTGCAAAATTGTTCGGCGACGTCCAGCACATCATCAA
GGAGCTGGACATACCAACAAGCTGGCGAAGCGCATCACCGAGACGCCGAAAG
ATTACGAGGCAGCGATCGAGTTACGCGAGAATGCGCATCTGAAGAAGAAGTTCA
AGCCGCAAGGTAGCATCGAGCACCTGCAGAGCGGCGTCTACTACCTGACGAACA
30 TTGACGACAAGTTCGCGCGTTCTTATGACGTCAAAAAGTAACTAGTAGGAGGAA
AACATCATGGTGCTGACGAACAAAACCGTCATTAGCGGCAGCAAGGTGAAGTCT
CTGAGCAGCGCCCAAAGCTCTAGCAGCGGCCCGTCTAGCAGCAGCGAGGAGGAC
35 GACAGCCGTGACATTGAGTCTCTGGACAAGAAGATCCGCCCCGTGGAGGAGTTA
GAGGCCCTGCTGAGCAGCGGCAACACCAAGCAGCTGAAGAACAAGGAAGTTGC
AGCGCTGGTGATCCACGGTAAGCTGCCACTGTATGCGCTGGAAAAGAACTGGG
40 CGATACGACGCGTGCGGTGCGGTGCGTCGCAAAGCCTTAAGCATCTTAGCGGA
GGCCCCGGTGTTAGCCAGCGACCGCCTGCCGTACAAGAACTACGACTACGACCG
CGTGTTTGGCGCGTGCTGCGAGAATGTCATTGGCTACATGCCGTTACCGGTTGGT
GTGATCGGCCCCGCTGGTCATTGATGGCACGAGCTATCACATTCCAATGGCGACCA
45 CGGAAGGTTGCTTAGTCGCCAGCGCCATGCGTGGCTGTAAGGCGATTAAACGCCG
GCGGTGGCGCGACGACCGTGTTAACCAAGGATGGTATGACGCGCGGTCCGGTCG
TCCGCTTCCCAACGCTGAAGCGCAGCGGCGCGTGTAAGATTTGGCTGGATTCTGA
50 GGAGGGCCAAAACGCGATCAAGAAAGCCTTCAACTCTACGAGCCGTTTCGCGCG

TTTACAGCATATCCAGACCTGCCTGGCCGGCGACCTGCTGTTTCATGCGCTTCCGC
 ACCACCACGGGCGATGCGATGGGCATGAACATGATCAGCAAGGGCGTCGAATAT
 5 AGCCTGAAACAAATGGTGGAGAATATGGCTGGGAGGACATGGAGGTTGTCTCT
 GTGAGCGGCAACTATTGCACCGACAAGAAGCCGGCAGCCATTAAGTGGATTGAG
 GGTCGCGGCAAAAGCGTCGTGGCAGAAGCGACCATCCCAGGCGACGTGGTCCGT
 10 AAGGTTCTGAAGAGCGACGTCAGCGCCCTGGTTGAGTTAAATATCGCGAAAAAC
 CTGGTTCGGCAGCGCGATGGCGGGCAGCGTGGGTGGCTTTAACGCACATGCAGCG
 AATCTGGTTACGGCGGTTTTCTTAGCCTTAGGTCAGGACCCAGCCCAAAATGTCG
 15 AGAGCAGCAACTGCATTACCTTAATGAAAGAGGTTGACGGTGACCTGCGCATCA
 GCGTTTCTATGCCGTCTATCGAGGTCGGCACGATCGGCGGGCGGCACCGTTTTAGA
 ACCGCAAGGTGCGATGCTGGATCTGCTGGGCGTGCGCGGCCACATGCAACGGC
 CCCAGGCACCAATGCCCCGCAACTGGCCCGTATCGTGGCCTGCGCGGTTCTGGCG
 20 GGTGAGCTGAGCCTGTGCGCCGCATTAGCCGCGGGCCATTTAGTTCAATCTCACA
 TGACCCACAACCGCAAGCCGGCAGAACCAACCAAGCCAAATAACCTGGACGCAA
 CCGACATTAACCGTCTGAAGGATGGCAGCGTCACGTGCATTAAGCTGAGCAT
 25 GCTACTAAGCTT

SEQ ID NO: 2

Primer 4-49 mvaA SpeI

5'-GCTACTAGTAGGAGGAAAACATCATGCAAAGTTTAGATAAGAATTTCCG-3'

SEQ ID NO: 3

Primer 4-49 mvaAR XbaI

5'-GCTTCTAGACTATTGTTGTCTAATTTCTGTAAAATGCG-3'

SEQ ID NO: 4

Primer HMGS 5' Sa mvaS-S

5'-

GAAGTGAAGATCTAGGAGGAAAGCAAAATGACAATAGGTATCGACAAAATAAA
 CT-3'

SEQ ID NO: 5

Primer HMGS 3' Sa mvaS-AS

5'-TTGCATGATGTTTTCTCCTACTAGTTACTCTGGTCTGTGATATTCGCGAAC-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'-CCACCTCGAGATGTCATTACCGTTCTTAACCTTG-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'-AGCTAAGCTTTTATTGTTTTCTTAAATCATTTAAATAGC - 3'

SEQ ID NO: 13

Primer 4-40 mvaSF BglII

5' – TATAGATCTTAAGGAGGATATTTAGATGACAATTGGGATTGATAAAATTAG –
3'

SEQ ID NO: 14

Primer 4-39 mvaSR BamHI

5'-TTTGGATCCTTAGTTTCGATAAGAGCGAACGG - 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 67-1 C-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 67-1 F-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'

5

SEQ ID NO: 22

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

5'- TOT 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'

10

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'

15

SEQ ID NO: 25

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

20

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

25

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'

30

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'

35

SEQ ID NO: 29

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

40

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

45

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'

50

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'

55

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 cm1-pAM37-S
5'- TACTAGTGCTTGGATTCTCACC - 3'

SEQ ID NO: 37

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

ATGAAGAACTGTGTGATTGTTTCTGCGGTCCGCACGGCGATCGGCAGCTTTAACG
GCTCTTTAGCGAGCACCTCTGCAATCGATCTGGGTGCGACGGTCATTAAGGCCGC
5 CATTGAACGCGCCAAAATCGACAGCCAGCACGTTGATGAGGTGATCATGGGCAA
TGTGTTACAAGCCGGCCTGGGTCAAAACCCAGCGCGTCAAGCACTGTTAAAATCT
GGTCTGGCCGAGACCGTGTGTGGCTTCACCGTCAATAAGGTTTGCGGCTCTGGCC
10 TGAAGAGCGTGGCCCTGGCAGCACAAGCGATTCAAGCCGGTCAGGCACAAAGCA
TCGTTGCGGGTGGCATGGAGAACATGTCTCTGGCGCCGTACTTATTAGATGCCAA
AGCCCGCAGCGGTTATCGCCTGGGCGATGGTCAGGTGTACGACGTCATCTTACGC
15 GATGGCTTAATGTGCGCGACCCACGGTTACCACATGGGTATTACGGCCGAAAAC
GTGGCGAAAGAATACGGCATTACGCGCGAGATGCAGGATGAATTAGCACTGCAC
TCTCAGCGCAAAGCAGCAGCCGCGATCGAGTCTGGTGCGTTTACGGCGGAAATC
GTGCCAGTTAACGTGGTCACGCGCAAGAAGACGTTTCGTTTTACGCCAGGACGAG
20 TTCCCGAAGGCAAACAGCACCGCGGAGGCCTTAGGTGCCTTACGCCAGCCTTTG
ACAAAGCGGGCACGGTCACCGCCGGTAATGCGAGCGGCATCAATGATGGTGCAG
CGGCACTGGTCATCATGGAAGAGAGCGCCGCATTAGCAGCGGGTCTGACCCCAT
25 TAGCGCGCATTAAATCTTATGCCAGCGGCGGCGTCCCACCAGCCCTGATGGGCAT
GGGTCCGGTCCCAGCCACGCAAAAAGCCCTGCAATTAGCGGGCCTGCAACTGGC
CGACATTGATCTGATCGAGGCGAACGAGGCGTTTGCAGCGCAGTTCCTGGCGGT
30 GGGTAAGAATCTGGGCTTCGACAGCGAGAAAGTCAATGTGAACGGTGGCGCGAT
TGCGTTAGGCCATCCGATTGGTGCAAGCGGCGCACGCATCTTAGTGACGTTACTG
CACGCCATGCAGGCACGCGACAAGACCTTAGGCCTGGCGACCTTATGTATTGGTG
GCGGTCAAGGTATCGCCATGGTGATCGAACGCCTGAACTGAAGATCTAGGAGGA
35 AAGCAAAATGAACTGAGCACCAAGCTGTGCTGGTGTGGCATCAAGGGTCGCCT
GCGCCCACAAAAGCAGCAACAGCTGCACAACACGAACCTGCAAATGACCGAGCT

GAAAAAGCAGAAGACGGCCGAGCAAAAGACCCGCCCGCAGAACGTTGGCATCA
 AGGGCATCCAGATTTATATCCCGACGCAGTGTGTCAACCAATCTGAGCTGGAGA
 5 AATTCGATGGCGTCAGCCAGGGTAAGTACACCATCGGCCTGGGCCAGACCAACA
 TGAGCTTCGTGAACGACCGTGAGGACATCTATTCTATGAGCCTGACGGTGCTGTC
 TAAGCTGATCAAGAGCTACAACATCGACACGAATAAGATCGGTCTGTCTGGAGGT
 10 GGGTACGGAGACGCTGATTGACAAGAGCAAAAGCGTGAAGTCTGTCTTAATGCA
 GCTGTTCCGGCGAGAACACGGATGTCGAGGGTATCGACACCCTGAACGCGTGTTA
 CGGCGGCACCAACGCACTGTTCAATAGCCTGAACTGGATTGAGAGCAACGCCTG
 15 GGATGGCCGCGATGCGATCGTCGTGTGCGGCGATATCGCCATCTATGACAAGGG
 TCGCGCACGTCCGACCGGCGGTGCAGGCACCGTTGCGATGTGGATTGGCCCGGA
 CGCACCAATTGTCTTCGATTCTGTCCGCGCGTCTTACATGGAGCACGCCTACGAC
 TTTTACAAGCCGGACTTCACGAGCGAATACCCGTACGTGGACGGCCACTTCTCTC
 20 TGACCTGCTATGTGAAGGCGCTGGACCAGGTTTATAAGTCTTATAGCAAAAAGGC
 GATTTCTAAGGGCCTGGTCAGCGACCCGGCAGGCAGCGACGCCCTGAACGTGCT
 GAAGTATTTGACTACAACGTGTTCCATGTCCCGACCTGCAAATTAGTGACCAAA
 25 TCTTATGGCCGCCTGTTATATAATGATTTCCGTGCCAACCCGCAGCTGTTCCCGGA
 GGTTGACGCCGAGCTGGCGACGCGTGATTACGACGAGAGCCTGACCGACAAGAA
 CATCGAGAAGACCTTCGTCAACGTCGCGAAGCCGTTCCACAAAGAGCGTGTGGC
 30 CCAAAGCCTGATCGTCCCGACCAACACGGGCAACATGTATACCGCGTCTGTCTAC
 GCGGCATTCCGCGAGCCTGCTGAATTACGTTCGGTTCTGACGACCTGCAGGGCAAGC
 GCGTTGGCCTGTTTCAGCTACGGTAGCGGCTTAGCGGCCAGCCTGTATAGCTGCAA
 35 AATTGTCGGCGACGTCCAGCACATCATCAAGGAGCTGGACATCACCAACAAGCT
 GGCGAAGCGCATCACCGAGACGCCGAAAGATTACGAGGCAGCGATCGAGTTACG
 CGAGAATGCGCATCTGAAGAAGAACTTCAAGCCGCAAGGTAGCATCGAGCACCT
 GCAGAGCGGCGTCTACTACCTGACGAACATTGACGACAAGTTCCGCCGTTCTTAT
 40 GACGTCAAAAAGTAACTAGTAGGAGGAAAACATCATGCAAAGTTTAGATAAGAA
 TTTCCGACATTTATCTCGTCAACAAAAGTTACAACAATTGGTAGATAAGCAATGG
 TTATCAGAAGATCAATTCGACATTTTATTGAATCATCCATTAATTGATGAGGAAG
 45 TAGCAAATAGTTTAATTGAAAATGTCATCGCGCAAGGTGCATTACCCGTTGGATT
 ATTACCGAATATCATTGTGGACGATAAGGCATATGTTGTACCTATGATGGTGGAA
 GAGCCTTCAGTTGTCGCTGCAGCTAGTTATGGTGCAAAGCTAGTGAATCAGACTG
 50 GCGGATTTAAAACGGTATCTTCTGAACGTATTATGATAGGTCAAATCGTCTTTGA
 TGGCGTTGACGATACTGAAAAATTATCAGCAGACATTAAAGCTTTAGAAAAGCA
 AATTCATAAAATTGCGGATGAGGCATATCCTTCTATTAAAGCGCGTGGTGGTGGT

TACCAACGTATAGCTATTGATACATTTCTGAGCAACAGTTACTATCTTTAAAAG
 TATTTGTTGATACGAAAGATGCTATGGGCGCTAATATGCTTAATACGATTTTAGA
 5 GGCCATAACTGCATTTTTTAAAAAATGAATCTCCACAAAGCGACATTTTAATGAGT
 ATTTTATCCAATCATGCAACAGCGTCCGTTGTTAAAGTTCAAGGCGAAATTGACG
 TTAAAGATTTAGCAAGGGGCGAGAGAACTGGAGAAGAGGTTGCCAAACGAATG
 10 GAACGTGCTTCTGTATTGGCACAAGTTGATATTCATCGTGCTGCAACACATAATA
 AAGGTGTTATGAATGGCATAACATGCCGTTGTTTTAGCAACAGGAAATGATACGCG
 TGGTGCAGAAGCAAGTGCGCATGCATACGCGAGTCGTGACGGACAGTATCGTGG
 TATTGCAACATGGAGATACGATCAAAAACGTCAACGTTTAATTGGTACAATAGA
 15 AGTGCCTATGACATTGGCAATCGTTGGCGGTGGTACAAAAGTATTACCAATTGCT
 AAAGCTTCTTTAGAATTGCTAAATGTAGATTCAGCACAAAGAATTAGGTCATGTAG
 TTGCTGCCGTTGGTTTAGCACAGAACTTTGCAGCATGTCGCGCGCTCGTTTCCGA
 20 AGGTATCCAGCAAGGCCATATGAGCTTGCAATATAAATCTTTAGCTATTGTTGTA
 GGTGCAAAAGGTGATGAAATTGCGCAAGTAGCTGAAGCATTGAAGCAAGAACCC
 CGTGCGAATACACAAGTAGCTGAACGCATTTTACAAGAAATTAGACAACAATAG

SEQ ID NO: 42

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

ATGAAGAACTGTGTGATTGTTTCTGCGGTCCGCACGGCGATCGGCAGCTTTAACG
 GCTCTTTAGCGAGCACCTCTGCAATCGATCTGGGTGCGACGGTCATTAAGGCCCGC
 CATTGAACGCGCCAAAATCGACAGCCAGCACGTTGATGAGGTGATCATGGGCAA
 35 TGTGTTACAAGCCGGCCTGGGTCAAAACCCAGCGCGTCAAGCACTGTTAAAATCT
 GGTCTGGCCGAGACCGTGTGTGGCTTCACCGTCAATAAGGTTTGCGGCTCTGGCC
 TGAAGAGCGTGGCCCTGGCAGCACAAAGCGATTCAAGCCGGTCAGGCACAAAGCA
 40 TCGTTGCGGGTGGCATGGAGAACATGTCTCTGGCGCCGTACTTATTAGATGCCAA
 AGCCCGCAGCGGTTATCGCCTGGGCGATGGTCAGGTGTACGACGTCATCTTACGC
 GATGGCTTAATGTGCGCGACCCACGGTTACCACATGGGTATTACGGCCGAAAAC
 GTGGCGAAAGAATACGGCATTACGCGCGAGATGCAGGATGAATTAGCACTGCAC
 45 TCTCAGCGCAAAGCAGCAGCCGCGATCGAGTCTGGTGCGTTTACGGCGGAAATC
 GTGCCAGTTAACGTGGTCACGCGCAAGAAGACGTTTCGTTTTACGCCAGGACGAG
 TTCCCGAAGGCAAACAGCACCGCGGAGGCCTTAGGTGCCTTACGCCAGCCTTTG
 50 ACAAAGCGGGCACGGTCACCGCCGGTAATGCGAGCGGCATCAATGATGGTGCA
 CGGCACTGGTCATCATGGAAGAGAGCGCCGCATTAGCAGCGGGTCTGACCCCAT
 TAGCGCGCATTAATCTTATGCCAGCGGCGGCGTCCCACCAGCCCTGATGGGCAT

GGGTCCGGTCCCAGCCACGCAAAAAGCCCTGCAATTAGCGGGCCTGCAACTGGC
CGACATTGATCTGATCGAGGCGAACGAGGCGTTTGCAGCGCAGTTCCTGGCGGT
GGGTAAGAATCTGGGCTTCGACAGCGAGAAAGTCAATGTGAACGGTGGCGCGAT
TGCGTTAGGCCATCCGATTGGTGCAAGCGGCGCACGCATCTTAGTGACGTTACTG
CACGCCATGCAGGCACGCGACAAGACCTTAGGCCTGGCGACCTTATGTATTGGTG
GCGGTCAAGGTATCGCCATGGTGATCGAACGCCTGAACTGAAGATCTAGGAGGA
AAGCAAAATGACAATAGGTATCGACAAAATAAACTTTTACGTTCCAAAGTACTA
TG TAGACATGGCTAAATTAGCAGAAGCACGCCAAGTAGACCCAAACAAATTTTT
AATTGGAATTGGTCAAACCTGAAATGGCTGTTAGTCCTGTAAACCAAGACATCGTT
TCAATGGGCGCTAACGCTGCTAAGGACATTATAACAGACGAAGATAAAAAGAAA
ATTGGTATGGTAATTGTGGCAACTGAATCAGCAGTTGATGCTGCTAAAGCAGCCG
CTGTTCAAATTCACAACCTTATTAGGTATTCAACCTTTTGCACGTTGCTTTGAAATG
AAAGAAGCTTGTTATGCTGCAACACCAGCAATTCAATTAGCTAAAGATTATTTAG
CAACTAGACCGAATGAAAAAGTATTAGTTATTGCTACAGATACAGCACGTTATG
GATTGAATTCAGGCGGCGAGCCAACACAAGGTGCTGGCGCAGTTGCGATGGTTA
TTGCACATAATCCAAGCATTTTGGCATTAAATGAAGATGCTGTTGCTTACACTGA
AGACGTTTATGATTTCTGGCGTCCAACCTGGACATAAATATCCATTAGTTGATGGT
GCATTATCTAAAGATGCTTATATCCGCTCATTCCAACAAAGCTGGAATGAATACG
CAAAACGTCAAGGTAAGTCGCTAGCTGACTTCGCATCTCTATGCTTCCATGTTCC
ATTTACAAAAATGGGTAAAAAGGCATTAGAGTCAATCATTGATAACGCTGATGA
AACAACTCAAGAGCGTTTACGTTCAAGGATATGAAGATGCTGTAGATTATAACCGT
TATGTCGGTAATATTTATACTGGATCATTATATTTAAGCCTAATATCATTACTTGA
AAATCGTGATTTACAAGCTGGTGAAACAATCGGTTTATTCAGTTATGGCTCAGGT
TCAGTTGGTGAATTTTATAGTGCGACATTAGTTGAAGGCTACAAAGATCATTTAG
ATCAAGCTGCACATAAAGCATTATTAATAACCGTACTGAAGTATCTGTTGATGC
ATATGAAACATTCTTCAAACGTTTTTGATGACGTTGAATTTGACGAAGAACAAGAT
GCTGTTTATGAAGATCGTCATATTTTCTACTTATCAAATATTGAAAATAACGTTCCG
CGAATATCACAGACCAGAGTAACTAGTAGGAGGAAAACATCATGCAAAGTTTAG
ATAAGAATTTCCGACATTTATCTCGTCAACAAAAGTTACAACAATTGGTAGATAA
GCAATGGTTATCAGAAGATCAATTCGACATTTTATTGAATCATCCATTAATTGAT
GAGGAAGTAGCAAATAGTTTAATTGAAAATGTCATCGCGCAAGGTGCATTACCC
GTTGGATTATTACCGAATATCATTGTGGACGATAAGGCATATGTTGTACCTATGA
TGGTGGAAGAGCCTTCAGTTGTCGCTGCAGCTAGTTATGGTGCAAAGCTAGTGAA
TCAGACTGGCGGATTTAAAACGGTATCTTCTGAACGTATTATGATAGGTCAAATC

GTCTTTGATGGCGTTGACGATACTGAAAAATTATCAGCAGACATTAAAGCTTTAG
 AAAAGCAAATTCATAAAATTGCGGATGAGGCATATCCTTCTATTAAAGCGCGTG
 5 GTGGTGGTTACCAACGTATAGCTATTGATACATTTCTGAGCAACAGTTACTATC
 TTTAAAAGTATTTGTTGATACGAAAGATGCTATGGGCGCTAATATGCTTAATACG
 ATTTTAGAGGCCATAACTGCATTTTAAAAAATGAATCTCCACAAAGCGACATTT
 10 TAATGAGTATTTTATCCAATCATGCAACAGCGTCCGTTGTTAAAGTTCAAGGCGA
 AATTGACGTTAAAGATTTAGCAAGGGGCGAGAGAACTGGAGAAGAGGTTGCCAA
 ACGAATGGAACGTGCTTCTGTATTGGCACAAGTTGATATTCATCGTGCTGCAACA
 15 CATAATAAAGGTGTTATGAATGGCATAACATGCCGTTGTTTTAGCAACAGGAAATG
 ATACGCGTGGTGCAGAAGCAAGTGCGCATGCATACGCGAGTCGTGACGGACAGT
 ATCGTGGTATTGCAACATGGAGATACGATCAAAAACGTCAACGTTTAATTGGTAC
 AATAGAAGTGCCTATGACATTGGCAATCGTTGGCGGTGGTACAAAAGTATTACC
 20 AATTGCTAAAGCTTCTTTAGAAATTGCTAAATGTAGATTCAGCACAGAATTAGGT
 CATGTAGTTGCTGCCGTTGGTTTAGCACAGAACTTTGCAGCATGTCGCGCGCTCG
 TTTCCGAAGGTATCCAGCAAGGCCATATGAGCTTGCAATATAAATCTTTAGCTAT
 25 TGTGTAGGTGCAAAAGGTGATGAAATTGCGCAAGTAGCTGAAGCATTGAAGCA
 AGAACCCCGTGCGAATACACAAGTAGCTGAACGCATTTTACAAGAAATTAGACA
 ACAATAG

SEQ ID NO: 43

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

35 CAATACCGACTTACCATCCTATTTGCTTTGCCCTTTTCTTTTCCACTGCATGGCG
 GCGTTAGTATCGAATGGATGGCGGCGTTAGTATCGAATCGACAGCAGTATAGCG
 ACCAGCATTACATACGATTGACGCATGATATTACTTTCTGCGCACTTAACCTCG
 40 CATCTGGGCAGATGATGTCGAGGCGAAAAAAATATAAATCACGCTAACATTTG
 ATTTAAATAGAACAACACTACAATATAAAAAAACTATACAAATGACAAGTTCTTGA
 AAACAAGAATCTTTTTATTGTCAGTACTGATTAGAAAACTCATCGAGCATCAAA
 TGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTGA AAAAGCC
 45 GTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGAT
 CCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTC
 CCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAA
 50 TCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCC
 AGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTG
 TGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAGGACAATTACA

AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT
TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTGCCGGGGATC
5 GCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTC
GGAAGAGGCATAAATTCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACAT
CATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTT
10 CCCATACAATCGATAGATTGTCGCACCTGATTGCCCCGACATTATCGCGAGCCCAT
TTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAAACGT
GAGTCTTTTCCTTACCCATGGTTGTTTATGTTCCGGATGTGATGTGAGAACTGTATC
CTAGCAAGATTTTAAAAGGAAGTATATGAAAGAAGAACCTCAGTGGCAAATCCT
15 AACCTTTTATATTTCTCTACAGGGGCGCGGCGTGGGGACAATTCAACGCGTCTGT
GAGGGGAGCGTTTCCCTGCTCGCAGGTCTGCAGCGAGGAGCCGTAATTTTGTCT
CGCGCCGTGCGGCCATCAAATGTATGGATGCAAATGATTATACATGGGGATGT
20 ATGGGCTAAATGTACGGGCGACAGTCACATCATGCCCCTGAGCTGCGCACGTCA
AGACTGTCAAGGAGGGTATTCTGGGCCTCCATGTCGCTGGCCGGGTGACCCGGC
GGGGACGAGGCAAGCTAAACAGATCTGATCTTGAAACTGAGTAAGATGCTCAGA
ATACCCGTCAAGATAAGAGTATAATGTAGAGTAATATACCAAGTATTCAGCATAT
25 TCTCCTCTTCTTTTGTATAAATCACGGAAGGGATGATTTATAAGAAAAATGAATA
CTATTACACTTCATTTACCACCCTCTGATCTAGATTTTCCAACGATATGTACGTAG
TGGTATAAGGTGAGGGGGTCCACAGATATAACATCGTTTAATTTAGTACTAACAG
30 AGACTTTTGTCACAACTACATATAAGTGTACAAATATAGTACAGATATGACACAC
TTGTAGCGCCAACGCGCATCCTACGGATTGCTGACAGAAAAAAGGTCACGTGA
CCAGAAAAGTCACGTGTAATTTTGTAACTCACCGCATTCTAGCGGTCCCTGTCGT
35 GCACACTGCACTCAACACCATAAACCTTAGCAACCTCCAAAGGAAATCACCGTA
TAACAAAGCCACAGTTTTACAACCTTAGTCTCTTATGAAGTTACTTACCAATGAGA
AATAGAGGCTCTTTCTCGAGAAATATGAATATGGATATATATATATATATATA
40 TATATATATATATATGTAACTTGGTTCTTTTTTAGCTTGTGATCTCTAGCTTGGG
TCTCTCTCTGTCGTAACAGTTGTGATATCGGCTGCCTTCATCTCGACCGGATGCAA
TGCCAATTGTAATAGCTTTCCCATGTTAATTATACTTTATTCTT

SEQ ID NO: 44

GAGTGAACCTGCTGCCTGGCGTGCTCTGACTCAGTACATTTCATAGTGGATGGCG
50 GCGTTAGTATC

SEQ ID NO: 45

CGTGTATACGTTTTCCGCTTCTGCTCTTCGTCTTTTCTCTTCTTCCGATATCACAAC
55 TGTTACGA

SEQ ID NO: 46

5 GGTAAGACGGTTGGGTTTTATCTTTGCAGTTGGTACTATTAAGAACAATCACAG
GAAACAGCTATGACC

SEQ ID NO: 47

10 TTGCGTTTTGTACTTTGGTTCGCTCAATTTTGCAGGTAGATAATCGAAAAGTTGTA
AAACGACGGCCAGT

SEQ ID NO: 48

pAM491 sequence (excluding vector backbone)

20 GTTTAAACTTGCTAAATTCGAGTGAAACACAGGAAGACCAGAAAATCCTCATTTC
ATCCATATTAACAATAATTTCAAATGTTTATTTGCATTATTTGAAACTAGGGAAG
ACAAGCAACGAAACGTTTTGAAAATTTTGAGTATTTTCAATAAATTTGTAGAGGA
25 CTCAGATATTGAAAAAAGCTACAGCAATTAATACTTGATAAGAAGAGTATTGA
GAAGGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGTC
AAACGACGTTGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATGA
TAACAAACCGAAGTTATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAGT
30 GATGATATTTCATAAATAATGTAATTCTATATATGTTAATTACCTTTTTTGCGAGG
CATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGAAGGTAAATGTGGCTGTG
GTTTCAGGGTCCATACCCGGGAGTTATGACAATTACAACAACAGAATTCTTTCTA
35 TATATGCACGAACTTGTAATATGGAAGAAATTATGACGTACAAACTATAAAGTA
AATATTTTACGTAACACATGGTGCTGTTGTGCTTCTTTTTCAAGAGAATACCAATG
ACGTATGACTAAGTTTAGGATTTAATGCAGGTGACGGACCCATCTTTCAAACGAT
40 TTATATCAGTGGCGTCCAAATTGTTAGGTTTTGTTGGTTCAGCAGGTTTCCTGTTG
TGGGTCATATGACTTTGAACCAAATGGCCGGCTGCTAGGGCAGCACATAAGGAT
AATTCACCTGCCAAGACGGCACAGGCAACTATTCTTGCTAATTGACGTGCGTTGG
45 TACCAGGAGCGGTAGCATGTGGGCCTCTTACACCTAATAAGTCCAACATGGCACC
TTGTGGTTCTAGAACAGTACCACCACCGATGGTACCTACTTCGATGGATGGCATG
GATACGGAAATTCTCAAATCACCGTCCACTTCTTTTCATCAATGTTATACAGTTGG
50 AACTTTCGACATTTTGTGCAGGATCTTGTCTTAATGCCAAGAAAACAGCTGTCAC

TAAATTAGCTGCATGTGCGTTAAATCCACCAACAGACCCAGCCATTGCAGATCCA
ACCAAATTCTTAGCAATGTTCAACTCAACCAATGCGGAAACATCACTTTTTAACA
5 CTTTTCTGACAACATCACCAGGAATAGTAGCTTCTGCGACGACACTCTTACCACG
ACCTTCGATCCAGTTGATGGCAGCTGGTTTTTTGTCGGTACAGTAGTTACCAGAA
ACGGAGACAACCTCCATATCTTCCCAGCCATACTCTTCTACCATTTGCTTTAATGA
10 GTATTCGACACCCTTAGAAATCATATTCATACCCATTGCGTCACCAGTAGTTGTTT
TAAATCTCATGAAGAGTAAATCTCCTGCTAGACAAGTTTGAATATGTTGCAGACG
TGCAAATCTTGATGTAGAGTTAAAAGCTTTTTTAATTGCGTTTTGTCCCTCTTCTG
15 AGTCTAACCATATCTTACAGGCACCAGATCTTTTCAAAGTTGGGAAACGGACTAC
TGGGCCTCTTGTCATACCATCCTTAGTTAAAACAGTTGTTGCACCACCGCCAGCA
TTGATTGCCTTACAGCCACGCATGGCAGAAGCTACCAAACAACCCTCTGTAGTTG
CCATTGGTATATGATAAGATGTACCATCGATAACCAAGGGGCCTATAACACCAA
20 CGGGCAAAGGCATGTAACCTATAACATTTTCACAACAAGCGCCAAATACGCGGT
CGTAGTCATAATTTTTATATGGTAAACGATCAGATGCTAATACAGGAGCTTCTGC
CAAAATTGAAAGAGCCTTCCTACGTACCGCAACCGCTCTCGTAGTATCACCTAAT
25 TTTTCTCCAAAGCGTACAAAGGTAACCTACCGTGAATAACCAAGGCAGCGACCT
CTTTGTTCTTCAATTGTTTTGTATTTCCACTACTTAATAATGCTTCTAATTCTTCTA
AAGGACGTATTTTCTTATCCAAGCTTTCAATATCGCGGGAATCATCTTCCTCACTA
30 GATGATGAAGGTCCTGATGAGCTCGATTGCGCAGATGATAAACTTTTGACTTTTCG
ATCCAGAAATGACTGTTTTATTGGTTAAAACCTGGTGTAGAAGCCTTTTGTACAGG
AGCAGTAAAAGACTTCTTGGTGACTTCAGTCTTCACCAATTGGTCTGCAGCCATT
35 ATAGTTTTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGGA
TACTTTTATGACATTTGAATAAGAAGTAATACAAACCGAAAAATGTTGAAAGTATT
AGTTAAAGTGGTTATGCAGCTTTTGCATTTATATATCTGTTAATAGATCAAAAAT
CATCGCTTCGCTGATTAATTACCCAGAAATAAGGCTAAAAAACTAATCGCATT
40 TTATCCTATGGTTGTTAATTTGATTCGTTGATTTGAAGGTTTGTGGGGCCAGGT
CTGCCAATTTTTCTCTTCATAACCATAAAAGCTAGTATTGTAGAATCTTTATTGT
TCGGAGCAGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAG
45 ACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAGGGCTGTCGCCCCGCTCGGCGG
CTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGAAAGTT
CCAAAGAGAAGGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAAC
50 ATATAAGTAAGATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGA
TTATTAAACTTCTTTGCGTCCATCCAAAAAAAAGTAAGAATTTTTGAAAATTCA
ATATAAATGAAACTCTCAACTAACTTTGTTGGTGTGGTATTAAAGGAAGACTTA
55

GGCCGCAAAAGCAACAACAATTACACAATACAACTTGCAAATGACTGAACTAA
AAAAACAAAAGACCGCTGAACAAAAAACAGACCTCAAAATGTCCGGTATTAAAG
5 GTATCCAAATTTACATCCCAACTCAATGTGTCAACCAATCTGAGCTAGAGAAATT
TGATGGCGTTTCTCAAGGTAAATACACAATTGGTCTGGGCCAAACCAACATGTCT
TTTGTCAATGACAGAGAAGATATCTACTCGATGTCCCTAACTGTTTTGTCTAAGTT
10 GATCAAGAGTTACAACATCGACACCAACAAAATTGGTAGATTAGAAGTCGGTAC
TGAAACTCTGATTGACAAGTCCAAGTCTGTCAAGTCTGTCTTGATGCAATTGTTT
GGTGA AACACTGACGTGCAAGGTATTGACACGCTTAATGCCTGTTACGGTGGTA
CCAACGCGTTGTTCAACTCTTTGAACTGGATTGAATCTAACGCATGGGATGGTAG
15 AGACGCCATTGTAGTTTGCGGTGATATTGCCATCTACGATAAGGGTGCCGCAAGA
CCAACCGGTGGTGCCGGTACTGTTGCTATGTGGATCGGTCCTGATGCTCCAATTG
TATTTGACTCTGTAAGAGCTTCTTACATGGAACACGCCTACGATTTTACAAGCC
20 AGATTTACACCAGCGAATATCCTTACGTGCGATGGTCAATTTTICATTAACCTGTTACG
TCAAGGCTCTTGATCAAGTTTACAAGAGTTATTCCAAGAAGGCTATTTCTAAAGG
GTTGGTTAGCGATCCCGCTGGTTCGGATGCTTTGAACGTTTTGAAATATTTGACT
25 ACAACGTTTTCCATGTTCCAACCTGTAAATTGGTCACAAAATCATACGGTAGATT
ACTATATAACGATTTTCAGAGCCAATCCTCAATTGTTCCCGAAGTTGACGCCGAA
TTAGCTACTCGCGATTATGACGAATCTTTAACCGATAAGAACATTGAAAAAACTT
30 TTGTTAATGTTGCTAAGCCATTCCACAAAGAGAGAGTTGCCCAATCTTTGATTGT
TCCAACAAACACAGGTAACATGTACACCGCATCTGTTTATGCCGCCTTTGCATCT
CTATTAAACTATGTTGGATCTGACGACTTACAAGGCAAGCGTGTTGGTTTATTTTC
35 TTACGGTTCCCGTTTAGCTGCATCTCTATATTCTTGCAAAATTGTTGGTGACGTCC
AACATATTATCAAGGAATTAGATATTACTAACAAATTAGCCAAGAGAATCACCG
AAACTCCAAAGGATTACGAAGCTGCCATCGAATTGAGAGAAAATGCCCATTTGA
AGAAGAACTTCAAACCTCAAGGTTCCATTGAGCATTGCAAAGTGGTGTTTACTA
40 CTTGACCAACATCGATGACAAATTTAGAAGATCTTACGATGTTAAAAAATAATCT
TCCCCCATCGATTGCATCTTGCTGAACCCCTTCATAAATGCTTTATTTTTTTGGC
AGCCTGCTTTTTTTTAGCTCTCATTTAATAGAGTAGTTTTTTAATCTATACTAGG
45 AAAACTCTTTATTTAATAACAATGATATATATATACCCGGGAAGCTTTTCAATTC
ATCTTTTTTTTTTTTGTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTTTTGATTC
GGTAATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGG
50 TATATATACGCATATGTGGTGTGGAAGAAACATGAAATTGCCCAGTATTCTTAAC
CCAACCTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTCGAAAGC
TACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTT
55

AATATCATGCACGAAAAGCAAACAACTTGTGTGCTTCATTGGATGTTTCGTACCA
 CCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAATTTGTTTACTAAA
 5 AACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTA
 AAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTG
 ACATTGGTAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAG
 10 AATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCG
 GTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGT
 TAGCAGAATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTA
 15 CTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAG
 AGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGT
 GGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGA
 20 TGTGGTCTCTACAGGATCTGACATTATTATTGTTGGGTTTAAAC

SEQ ID NO: 49

pAM492 sequence (excluding vector backbone)

GTTTAAACTTGCTAAATTCGAGTGAAACACAGGAAGACCAGAAAATCCTCATTTC
 ATCCATATTAACAATAATTTCAAATGTTTATTTGCATTATTTGAAACTAGGGAAG
 30 ACAAGCAACGAAACGTTTTTGAAAATTTTGAGTATTTTCAATAAATTTGTAGAGG
 ACTCAGATATTGAAAAAAGCTACAGCAATTAATACTTGATAAGAAGAGTATTG
 AGAAGGGCAACGGTTCATCATCTCATGGATCTGCACATGAACAAACACCAGAGT
 CAAACGACGTTGAAATTGAGGCTACTGCGCCAATTGATGACAATACAGACGATG
 35 ATAACAAACCGAAGTTATCTGATGTAGAAAAGGATTAAAGATGCTAAGAGATAG
 TGATGATATTTCATAAATAATGTAATTCTATATATGTTAATTACCTTTTTTGCGAG
 GCATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGAAGGTTAATGTGGCTGT
 40 GGTTCAGGGTCCATACCCGGGTATATATATATCATTGTTATTAAATAAAGAGTT
 TTCCTAGTATATAGATTAAAAAACTACTCTATTAAATGAGAGCTAAAAAAAGCA
 GGCTGCCAAAAAATAAAGCATTATGAAGGGGGTTCAGCAAGATGCAATCGAT
 45 GGGGGAAGATTATTTTTTAACATCGTAAGATCTTCTAAATTTGTCATCGATGTTG
 GTCAAGTAGTAAACACCACTTTGCAAATGCTCAATGGAACCTTGAGGTTTGAAGT
 TCTTCTTCAAATGGGCATTTTCTCTCAATTTCGATGGCAGCTTCGTAATCCTTTGGA
 50 GTTTCGGTGATTCTCTTGGCTAATTTGTTAGTAATATCTAATTCCTTGATAATATG
 TTGGACGTCACCAACAATTTTGCAAGAATATAGAGATGCAGCTAAACCGGAACC
 GTAAGAAAATAAACCAACACGCTTGCCTTGTAAGTCGTCAGATCCAACATAGTTT
 55 AATAGAGATGCAAAGGCGGCATAAACAGATGCGGTGTACATGTTACCTGTGTTT

GTTGGAACAATCAAAGATTGGGCAACTCTCTCTTTGTGGAATGGCTTAGCAACAT
TAACAAAAGTTTTTCAATGTTCTTATCGGTAAAGATTTCGTCATAATCGCGAGT
5 AGCTAATTCGGCGTCAACTTCTGGGAACAATTGAGGATTGGCTCTGAAATCGTTA
TATAGTAATCTACCGTATGATTTTGTGACCAATTTACAGGTTGGAACATGGAAAA
CGTTGTAGTCGAAATATTTCAAACGTTCAAAGCATCCGAACCAGCGGGATCGCT
10 AACCAACCCTTTAGAAATAGCCTTCTTGGAATAACTCTTGTAAACTTGATCAAGA
GCCTTGACGTAAACAAGTTAATGAAAAATGACCATCGACGTAAGGATATTCGCTG
GTGAAATCTGGCTTGTA AAAATCGTAGGCGTGTTCCATGTAAGAAGCTCTTACAG
15 AGTCAAATACAATTGGAGCATCAGGACCGATCCACATAGCAACAGTACCGGCAC
CACCGGTTGGTCTTGCGGCACCTTATCGTAGATGGCAATATCACCGCAAACCTAC
AATGGCGTCTCTACCATCCCATGCGTTAGATTCAATCCAGTTCAAAGAGTTGAAC
AACGCGTTGGTACCACCGTAACAGGCATTAAGCGTGTCAATACCTTCGACGTCAG
20 TGTTTTACCAAACAATTGCATCAAGACAGACTTGACAGACTTGGACTTGTCAAT
CAGAGTTTCAGTACCGACTTCTAATCTACCAATTTTGTGTTGGTGTGCGATGTTGTAAC
TCTTGATCAACTTAGACAAAACAGTTAGGGACATCGAGTAGATATCTTCTCTGTC
25 ATTGACAAAAGACATGTTGGTTTGGCCCAGACCAATTGTGTATTTACCTTGAGAA
ACGCCATCAAATTTCTCTAGCTCAGATTGGTTGACACATTGAGTTGGGATGTAAA
TTTGGATACCTTTAATACCGACATTTTGAGGTCTGGTTTTTTGTTCAGCGGTCTTT
30 TGTTTTTTTAGTTCAAGTCTTGTCAAGTTTGTATTGTGTAATTGTTGTTGCTTTTGC
GGCCTAAGTCTTCCTTTAATACCACACCAACAAAGTTTAGTTGAGAGTTTCATTTT
ATGTGATGATTGATTGATTGATTGTACAGTTTGTTTTTCTTAATATCTATTTTCGAT
35 GACTTCTATATGATATTGCACTAACAAGAAGATATTATAATGCAATTGATACAAG
ACAAGGAGTTATTTGCTTCTCTTTTATATGATTCTGACAATCCATATTGCGTTGGT
AGTCTTTTTTGCTGGAACGGTTCAGCGGAAAAGACGCATCGCTCTTTTTGCTTCTA
40 GAAGAAATGCCAGCAAAAAGAATCTCTTGACAGTGACTGACAGCAAAAATGTCTT
TTTCTAACTAGTAACAAGGCTAAGATATCAGCCTGAAATAAAGGGTGGTGAAGT
AATAATTAAATCATCCGTATAAACCTATACACATATATGAGGAAAAATAATACA
AAAGTGTTTTAAATACAGATACATACATGAACATATGCACGTATAGCGCCCAAAT
45 GTCGGTAATGGGATCGGCTTACTAATTATAAAATGCATCATAGAAATCGTTGAAG
TTGACGCAGCGACTCGAGATCCATAGGAGCAACTCATGTCTGAACTTCAACGATT
TCTATGATGCATTTTATAATTAGTAAGCCGATCCCATTACCGACATTTGGGCGCT
50 ATACGTGCATATGTTTCATGTATGTATCTGTATTTAAAACACTTTTGTATTATTTT
CCTCATATATGTGTATAGGTTTATACGGATGATTTAATTATTACTTCACCACCCTT
TATTTACAGGCTGATATCTTAGCCTTGTTACTAGTTAGAAAAAGACATTTTGTCTGT
55

CAGTCACTGTCAAGAGATTCTTTTGCTGGCATTCTTCTAGAAGCAAAAAGAGCG
 ATGCGTCTTTTCCGCTGAACCGTTCCAGCAAAAAGACTACCAACGCAATATGGA
 5 TTGTCAGAATCATATAAAAGAGAAGCAAATAACTCCTTGTCCTTGTATCAATTGCA
 TTATAATATCTTCTTGTTAGTGCAATATCATATAGAAGTCATCGAAATAGATATT
 AAGAAAAACAACTGTACAATCAATCAATCAATCATCACATAAAATGGCTGCAG
 10 ACCAATTGGTGAAGACTGAAGTCACCAAGAAGTCTTTTACTGCTCCTGTACAAAA
 GGCTTCTACACCAGTTTTAAACCAATAAAACAGTCATTTCTGGATCGAAAGTCAAA
 AGTTTATCATCTGCGCAATCGAGCTCATCAGGACCTTCATCATCTAGTGAGGAAG
 15 ATGATTCCCGCGATATTGAAAGCTTGGATAAGAAAATACGTCCTTTAGAAGAATT
 AGAAGCATTATTAAGTAGTGGAATACAAAACAATTGAAGAACAAAGAGGTGCG
 TGCCTTGTTATTACCGGTAAGTTACCTTTGTACGCTTTGGAGAAAAAATTAGGT
 20 GATACTACGAGAGCGGTTGCGGTACGTAGGAAGGCTCTTTCAATTTTGGCAGAA
 GCTCCTGTATTAGCATCTGATCGTTTACCATATAAAAATTATGACTACGACCGCG
 TATTTGGCGCTTGTTGTGAAAATGTTATAGGTTACATGCCTTTGCCCGTTGGTGTT
 ATAGGCCCCCTTGGTTATCGATGGTACATCTTATCATATACCAATGGCAACTACAG
 25 AGGGTTGTTTGGTAGCTTCTGCCATGCGTGGCTGTAAGGCAATCAATGCTGGCGG
 TGGTGCAACAACTGTTTTAACTAAGGATGGTATGACAAGAGGGCCAGTAGTCCGT
 TTCCCAACTTTGAAAAGATCTGGTGCCTGTAAGATATGGTTAGACTCAGAAGAGG
 30 GACAAAACGCAATTAaaaaaAGCTTTTAACTCTACATCAAGATTTGCACGTCTGCA
 ACATATTCAAACCTTGTCTAGCAGGAGATTTACTCTTCATGAGATTTAGAACAAC
 ACTGGTGACGCAATGGGTATGAATATGATTTCTAAGGGTGTGCAATACTCATTAA
 35 AGCAAATGGTAGAAGAGTATGGCTGGGAAGATATGGAGGTTGTCTCCGTTTCTG
 GTAACACTGTACCGACAAAAAACAGCTGCCATCAACTGGATCGAAGGTCGTG
 GTAAGAGTGTCGTCGCAGAAGCTACTATTCCTGGTGATGTTGTCAGAAAAGTGTT
 AAAAAGTGATGTTTCCGCATTGGTTGAGTTGAACATTGCTAAGAATTTGGTTGGA
 40 TCTGCAATGGCTGGGTCTGTTGGTGGATTTAACGCACATGCAGCTAATTTAGTGA
 CAGCTGTTTTCTTGGCATTAGGACAAGATCCTGCACAAAATGTCGAAAGTTCCAA
 CTGTATAACATTGATGAAAGAAGTGGACGGTGATTTGAGAATTTCCGTATCCATG
 45 CCATCCATCGAAGTAGGTACCATCGGTGGTGGTACTGTTCTAGAACCACAAGGTG
 CCATGTTGGACTTATTAGGTGTAAGAGGCCCCACATGCTACCGCTCCTGGTACCAA
 CGCACGTCAATTAGCAAGAATAGTTGCCTGTGCCGTCTTGGCAGGTGAATTATCC
 50 TTATGTGCTGCCCTAGCAGCCGGCCATTTGGTTCAAAGTCATATGACCCACAACA
 GGAAACCTGCTGAACCAACAAAACCTAACAATTTGGACGCCACTGATATAAATC
 GTTTGAAAGATGGGTCCGTCACCTGCATTAAATCCTAACTTAGTCATACGTCAT
 55

TGGTATTCTCTTGAAAAAGAAGCACAAACAGCACCATGTGTTACGTAAAATATTTA
 CTTTATAGTTTGTACGTCATAATTTCTTCCATATTACAAGTTCGTGCATATATAGA
 5 AAGAATTCTGTTGTTGTAATTGTCATAACTCCCGGGAAGCTTTTCAATTCATCTTT
 TTTTTTTTTGTTCTTTTTTTTGATTCCGGTTTCTTTGAAATTTTTTTGATTCCGGTAAT
 CTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATAT
 10 ACGCATATGTGGTGTGTAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTG
 CACAGAACAAAAACCTGCAGGAAACGAAGATAAATCATGTGCGAAAGCTACATAT
 AAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCA
 TGCACGAAAAGCAAACAACTTGTGTGCTTCATTGGATGTTTCGTACCACCAAGGA
 15 ATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAATTTGTTTACTAAAAACACAT
 GTGGATATCTTGACTGATTTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCAT
 TATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGG
 20 TAATACAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGC
 AGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAA
 GCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGA
 25 ATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGAC
 ATTGCGAAGAGCGACAAAGATTTTGTTATCGGCTTTATTGCTCAAAGAGACATGG
 GTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAG
 30 ATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCT
 CTACAGGATCTGACATTATTATTGTTGGGTTTAAAC

SEQ ID NO: 50

pAM489 sequence (excluding vector backbone)

GTTTAAACTACTATTAGCTGAATTGCCACTGCTATCGTTGTTAGTGGCGTTAGTGC
 TTGCATTCAAAGACATGGAGGGCGTTATTACGCCGGAGCTCCTCGACAGCAGATC
 40 TGATGACTGGTCAATATATTTTTGCATTGAGGCTCTGTTTGGAATTATATTTTGA
 ATGACCCATCTAATGTACTGGTATCACCAGATTTTCATGTCGTTTTTTAAAGCGGCT
 GCTTGAGTCTTAGCAATAGCGTCACCATCTGGTGAATCCTTTGAAGGAACCACTG
 45 ACGAAGGTTTGGACAGTGACGAAGAGGATCTTTCCTGCTTTGAATTAGTCGCGCT
 GGGAGCAGATGACGAGTTGGTGGAGCTGGGGGCAGGATTGCTGGCCGTCGTGGG
 TCCTGAATGGGTCTTGGCTGGTCCATCTCTATTCTGAAAACGGAAGAGGAGTAG
 50 GGAATATTACTGGCTGAAAATAAGTCTTGAATGAACGTATACGCGTATATTTCTA
 CCAATCTCTCAACACTGAGTAATGGTAGTTATAAGAAAGAGACCGAGTTAGGGA
 CAGTTAGAGGCGGTGGAGATATTCCTTATGGCATGTCTGGCGATGATAAACTTT

TCAAACGGCAGCCCCGATCTAAAAGAGCTGACACCCGGGAGTTATGACAATTAC
AACAAACAGAATTCTTTCTATATATGCACGAACTTGTAATATGGAAGAAATTATGA
5 CGTACAAACTATAAAGTAAATATTTTACGTAACACATGGTGCTGTTGTGCTTCTTT
TTCAAGAGAATACCAATGACGTATGACTAAGTTTAGGATTTAATGCAGGTGACG
GACCCATCTTTCAAACGATTTATATCAGTGGCGTCCAAATTGTTAGGTTTTGTTGG
10 TTCAGCAGGTTTTCTGTTGTGGGTCATATGACTTTGAACCAAATGGCCGGCTGCT
AGGGCAGCACATAAGGATAATTCACCTGCCAAGACGGCACAGGCAACTATTCTT
GCTAATTGACGTGCGTTGGTACCAGGAGCGGTAGCATGTGGGCCTCTTACACCTA
ATAAGTCCAACATGGCACCTTGTGGTTCTAGAACAGTACCACCACCGATGGTACC
15 TACTTCGATGGATGGCATGGATACGGAAATTCTCAAATCACCGTCCACTTCTTTC
ATCAATGTTATACAGTTGGAACCTTTCGACATTTTGTGCAGGATCTTGTCTAATGC
CAAGAAAACAGCTGTCACTAAATTAGCTGCATGTGCGTTAAATCCACCAACAGA
20 CCCAGCCATTGCAGATCCAACCAAATTCTTAGCAATGTTCAACTCAACCAATGCG
GAAACATCACTTTTTAACACTTTTCTGACAACATCACCAGGAATAGTAGCTTCTG
CGACGACACTCTTACCACGACCTTCGATCCAGTTGATGGCAGCTGGTTTTTTGTC
GGTACAGTAGTTACCAGAAACGGAGACAACCTCCATATCTTCCCAGCCATACTCT
25 TCTACCATTTGCTTTAATGAGTATTCGACACCCTTAGAAATCATATTCATACCCAT
TGCGTCACCAGTAGTTGTTCTAAATCTCATGAAGAGTAAATCTCCTGCTAGACAA
GTTTGAATATGTTGCAGACGTGCAAATCTTGATGTAGAGTTAAAAGCTTTTTTAA
30 TTGCGTTTTGTCCCTCTTCTGAGTCTAACCATATCTTACAGGCACCAGATCTTTTC
AAAGTTGGGAAACGGACTACTGGGCCTCTTGTGCATACCATCCTTAGTTAAAACAG
TTGTTGCACCACCGCCAGCATTGATTGCCTTACAGCCACGCATGGCAGAAGCTAC
35 CAAACAACCCTCTGTAGTTGCCATTGGTATATGATAAGATGTACCATCGATAACC
AAGGGGCCTATAACACCAACGGGCAAAGGCATGTAACTATAACATTTTCACAA
CAAGCGCCAAATACGCGGTCGTAGTCATAATTTTTATATGGTAAACGATCAGATG
40 CTAATACAGGAGCTTCTGCCAAAATTGAAAGAGCCTTCCTACGTACCGCAACCGC
TCTCGTAGTATCACCTAATTTTTTCTCCAAAGCGTACAAAGGTAACCTACCGTGA
ATAACCAAGGCAGCGACCTCTTTGTTCTTCAATTGTTTTGTATTCCACTACTTAA
TAATGCTTCTAATTCTTCTAAAGGACGTATTTTCTTATCCAAGCTTTCAATATCGC
45 GGGAATCATCTTCCTCACTAGATGATGAAGGTCCTGATGAGCTCGATTGCGCAGA
TGATAAACTTTTGACTTTCGATCCAGAAATGACTGTTTTATTGGTTAAAACCTGGTG
TAGAAGCCTTTTGTACAGGAGCAGTAAAAGACTTCTTGGTGACTTCAGTCTTCAC
50 CAATTGGTCTGCAGCCATTATAGTTTTTCTCCTTGACGTAAAGTATAGAGGTAT
ATTAACAATTTTTTGTGATACTTTTATGACATTTGAATAAGAAGTAATACAAAC

CGAAAATGTTGAAAGTATTAGTTAAAGTGGTTATGCAGCTTTTGCATTTATATAT
CTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCAGAAATAAGGC
5 TAAAAAACTAATCGCATTATTATCCTATGGTTGTAAATTTGATTTCGTTGATTTGAA
GGTTTGTGGGGCCAGGTTACTGCCAATTTTTCTCTTCATAACCATAAAAGCTAG
TATTGTAGAATCTTTATTGTTTCGGAGCAGTGC GGCGAGGCACATCTGCGTTTC
10 AGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTCGGAG
GGCTGTCGCCCCGCTCGGCGGCTTCTAATCCGTA CTTC AATATAGCAATGAGCAGT
TAAGCGTATTACTGAAAGTTCCAAAGAGAAGGTTTTTTTTAGGCTAAGATAATGGG
15 GCTCTTTACATTTCCACAACATATAAGTAAGATTAGATATGGATATGTATATGGT
GGTATTGCCATGTAATATGATTATTAACTTCTTTGCGTCCATCCAAAAAAAAG
TAAGAATTTTTGAAAATTCAATATAAATGGCTTCAGAAAAAGAAATTAGGAGAG
AGAGATTCTTGAACGTTTTCCCTAAATTAGTAGAGGAATTGAACGCATCGCTTTT
20 GGCTTACGGTATGCCTAAGGAAGCATGTGACTGGTATGCCCACTCATTGAACTAC
AACACTCCAGGCGGTAAGCTAAATAGAGGTTTGTCCGTTGTGGACACGTATGCTA
TTCTCTCCAACAAGACCGTTGAACAATTGGGGCAAGAAGAATACGAAAAGGTTG
25 CCATTCTAGGTTGGTGCATTGAGTTGTTGCAGGCTTACTTCTTGGTCGCCGATGAT
ATGATGGACAAGTCCATTACCAGAAGAGGCCAACCATGTTGGTACAAGGTTCTCT
GAAGTTGGGGAAATTGCCATCAATGACGCATTCATGTTAGAGGCTGCTATCTACA
30 AGCTTTTGAAATCTCACTTCAGAAACGAAAAATACTACATAGATATCACCGAATT
GTTCCATGAGGTCACCTTCCAAACCGAATTGGGCCAATTGATGGACTTAATCACT
GCACCTGAAGACAAAGTCGACTTGAGTAAGTTCTCCCTAAAGAAGCACTCCTTCA
TAGTTACTTTCAAGACTGCTTACTATTCTTTCTACTTGCCTGTTCGATTGGCCATG
35 TACGTTGCCGGTATCACGGATGAAAAGGATTTGAAACAAGCCAGAGATGTCTTG
ATTCCATTGGGTGAATACTTCCAAATTCAAGATGACTACTTAGACTGCTTCGGTA
CCCCAGAACAGATCGGTAAGATCGGTACAGATATCCAAGATAACAAATGTTCTT
40 GGGTAATCAACAAGGCATTGGAACCTTGCTTCCGCAGAACAAAGAAAGACTTTAG
ACGAAAATTACGGTAAGAAGGACTCAGTCGCAGAAAGCCAAATGCAAAAAGATTT
TCAATGACTTGAAAATTGAACAGCTATACCACGAATATGAAGAGTCTATTGCCAA
45 GGATTTGAAGGCCAAAATTTCTCAGGTCGATGAGTCTCGTGGCTTCAAAGCTGAT
GTCTTAACTGCGTTCTTGAACAAAGTTTACAAGAGAAGCAAATAGAACTAACGCT
AATCGATAAAACATTAGATTTCAAAC TAGATAAGGACCATGTATAAGAACTATA
50 TACTTCCAATATAATATAGTATAAGCTTTAAGATAGTATCTCTCGATCTACCGTTC
CACGTGACTAGTCCAAGGATTTTTTTTTAACCCGGGATATATGTGTACTTTGCAGTT
ATGACGCCAGATGGCAGTAGTGGAAGATATTCTTTATTGAAAAATAGCTTGTAC

CTTACGTACAATCTTGATCCGGAGCTTTTCTTTTTTTGCCGATTAAGAATTCGGTC
 GAAAAAAGAAAAGGAGAGGGCCAAGAGGGAGGGCATTGGTGACTATTGAGCAC
 5 GTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTATGTCTGTTATTAAT
 TTCACAGGTAGTTCTGGTCCATTGGTGAAAGTTTGCGGCTTGCAGAGCACAGAGG
 CCGCAGAATGTGCTCTAGATTCCGATGCTGACTTGCTGGGTATTATATGTGTGCC
 10 CAATAGAAAGAGAACAATTGACCCGGTTATTGCAAGGAAAATTTCAAGTCTTGT
 AAAAGCATATAAAAATAGTTCAGGCACTCCGAAATACTTGGTTGGCGTGTTCGT
 AATCAACCTAAGGAGGATGTTTTGGCTCTGGTCAATGATTACGGCATTGATATCG
 15 TCCAACCTGCATGGAGATGAGTCGTGGCAAGAATACCAAGAGTTCCTCGGTTTGCC
 AGTTATTAAAAGACTCGTATTTCCAAAAGACTGCAACATACTACTCAGTGCAGCT
 TCACAGAAACCTCATTCGTTTATTCCCTTGTTTGATTGAGAAGCAGGTGGGACAG
 20 GTGAACTTTTGGATTGGAACCTCGATTTCTGACTGGGTGGAAGGCAAGAGAGCCC
 CGAAAGCTTACATTTTATGTTAGCTGGTGGACTGACGCCGTTTAAAC

SEQ ID NO: 51

pAM497 sequence (excluding vector backbone)

GTTTAACTTTTCCAATAGGTGGTTAGCAATCGTCTTACTTTCTAACTTTTCTTAC
 30 CTTTACATTTCAAGCAATATATATATATATATTTCAAGGATATACCATTCTAATGT
 CTGCCCCTAAGAAGATCGTCGTTTTGCCAGGTGACCACGTTGGTCAAGAAATCAC
 AGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTTCGTTCCAATGTCAAG
 35 TTCGATTTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTTCC
 ACTTCCAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGT
 GCTGTGGGTGGTCCTAAATGGGGTACCGGTAGTGTTAGACCTGAACAAGGTTTAC
 TAAAAATCCGTAAAGAACTTCAATTGTACGCCAACTTAAGACCATGTAACCTTGC
 40 ATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTG
 ACTTCGTTGTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGA
 AGACGTTTAGCTTGCCTCGTCCCCGCGGGTCACCCGGCCAGCGACATGGAGGCC
 45 CAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACT
 GTCGCCCCTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACAT
 TTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGC
 50 GAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCC
 CTGTAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATAC
 TTCCTTTTAAAATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACA
 55 ACCATGGCAGAACCCAGCCCAAAAAAAGCAAAAACAACTGTTTCAGGAGCGCAA

GGCGTTTATCTCCCGTATCACTAATGAAACTAAAATTCAAATCGCTATTTTCGCTG
 AATGGTGGTTATATTCAAATAAAAAGATTTCGATTCTTCCTGCAAAGAAGGATGACG
 5 ATGTAGCTTCCCAAGCTACTCAGTCACAGGTCATCGATATTCACACAGGTGTTGG
 CTTTTTGGATCATATGATCCATGCGTTGGCAAAACACTCTGGTTGGTCTCTTATTG
 TTGAATGTATTGGTGACCTGCACATTGACGATCACCATACTACCGAAGATTGCGG
 10 TATCGCATTAGGGCAAGCGTTCAAAGAAGCAATGGGTGCTGTCCGTGGTGTA
 AAGATTTCGGTACTGGGTTCGCACCATTGGATGAGGCGCTATCACGTGCCGTAGTC
 GATTTATCTAGTAGACCATTTGCTGTAATCGACCTTGGATTGAAGAGAGAGATGA
 TTGGTGATTTATCCACTGAAATGATTCCACACTTTTTTGGAAAGTTTCGCGGAGGC
 15 GGCCAGAATTACTTTGCATGTTGATTGTCTGAGAGGTTTCAACGATCACCACAGA
 AGTGAGAGTGCGTTCAAGGCTTTGGCTGTTGCCATAAGAGAAGCTATTTCTAGCA
 ATGGCACCAATGACGTTCCCTCAACCAAAGGTGTTTTGATGTGAAGTACTGACAA
 20 TAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAG
 TTGTTCTATTTTAATCAAATGTTAGCGTGATTATATTTTTTTTCGCCTCGACATCA
 TCTGCCCAGATGCGAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTAT
 25 GTGAATGCTGGTCGCTATACTGCTGTCGATTTCGATACTAACGCCGCCATCCACCC
 GGGTTTCTCATTCAAGTGGTAACTGCTGTAAAATTAAGATATTTATAAATTGAA
 GCTTGGTCGTTCCGACCAATACCGTAGGGAAACGTAAATTAGCTATTGTAAAAAA
 30 AGGAAAAGAAAAGAAAAGAAAAATGTTACATATCGAATTGATCTTATTCTTTG
 GTAGACCAGTCTTTGCGTCAATCAAAGATTTCGTTTGTCTTGTGGGCCTGAACC
 GACTTGAGTTAAAATCACTCTGGCAACATCCTTTTGCAACTCAAGATCCAATTCA
 35 CGTGCAGTAAAGTTAGATGATTCAAATTGATGGTTGAAAGCCTCAAGCTGCTCAG
 TAGTAAATTTCTTGTTCCCATCCAGGAACAGAGCCAAACAATTTATAGATAAATGC
 AAAGAGTTTCGACTCATTTTCAGCTAAGTAGTACAACACAGCATTGACCTGCA
 TCAAACGTGTATGCAACGATTGTTTCTCCGTAAAACTGATTAATGGTGTGGCACC
 40 AACTGATGATACGCTTGGAAGTGTCATTCATGTAGAATATTGGAGGGAAAGAGT
 CCAAACATGTGGCATGGAAAGAGTTGGAATCCATCATTGTTTCTTTGCAAAGGT
 GGCGAAATCTTTTTCAACAATGGCTTTACGCATGACTTCAAATCTCTTTGGTACG
 45 ACATGTTCAATTCTTTCTTTAAATAGTTTCGGAGGTTGCCACGGTCAATTGCATACC
 CTGAGTGGAACCTCACATCCTTTTTAATATCGCTGACAACTAGGACACAAGCTTTC
 ATCTGAGGCCAGTCAGAGCTGTCTGCGATTTGTACTGCCATGGAATCATGACCAT
 50 CTTTCAGCTTTTCCCATTTCCAGGCCACGTATCCGCCAAACAACGATCTACAAGC
 TGAACCAGACCCCTTTCTTGCTATTCTAGATATTTCTGAAGTTGACTGTGGTAATT
 GGTATAACTTAGCAATTGCAGAGACCAATGCAGCAAAGCCAGCAGCGGAGGAAG
 55

CTAAACCAGCTGCTGTAGGAAAGTTATTTTCGGAGACAATGTGGAGTTTCCATTG
 AGATAATGTGGGCAATGAGGCGTCCTTCGATTCCATTTCTTTCTTAATTGGCGTA
 5 GGTGCGCGCAGACAATTTTGAGTTCTTTTCATTGTCGATGCTGTGTGGTTCTCCATTT
 AACCACAAAGTGTCGCGTTCAAACCTCAGGTGCAGTAGCCGCAGAGGTCAACGTT
 CTGAGGTCATCTTGCGATAAAGTCACTGATATGGACGAATTGGTGGGCAGATTCA
 10 ACTTCGTGTCCCTTTTCCCCCAATACTTAAGGGTTGCGATGTTGACGGGTGCGGT
 AACGGATGCTGTGTAAACGGTCATTATAGTTTTTTCTCCTTGACGTTAAAGTATA
 GAGGTATATTAACAATTTTTTGTGATACTTTTATGACATTTGAATAAGAAGTAAT
 15 ACAAACCGAAAAATGTTGAAAGTATTAGTTAAAGTGTTTATGCAGCTTTTGCATTT
 ATATATCTGTTAATAGATCAAAAATCATCGCTTCGCTGATTAATTACCCAGAAA
 TAAGGCTAAAAAACTAATCGCATTATTATCCTATGGTTGTTAATTTGATTCGTTGA
 TTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTTTTCTCCTTCATAACCATAAAA
 20 GCTAGTATTGTAGAATCTTTATTGTTTCGGAGCAGTGCGGCGCGAGGCACATCTGC
 GTTTCAGGAACGCGACCGGTGAAGACCAGGACGCACGGAGGAGAGTCTTCCGTC
 GGAGGGCTGTGCCCCGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAG
 25 CAGTTAAGCGTATTACTGAAAGTTCCAAAGAGAAGGTTTTTTTAGGCTAAGATAA
 TGGGGCTCTTTACATTTCCACAACATATAAGTAAGATTAGATATGGATATGTATA
 TGGTGGTATTGCCATGTAATATGATTATTAAACTTCTTTGCGTCCATCCAAAAAA
 30 AAAGTAAGAATTTTTGAAAATTCAATATAAATGTCAGAGTTGAGAGCCTTCAGTG
 CCCCAGGGAAAGCGTTACTAGCTGGTGGATATTTAGTTTTAGATCCGAAATATGA
 AGCATTGTAGTCGGATTATCGGCAAGAAATGCATGCTGTAGCCCATCCTTACGGT
 TCATTGCAAGAGTCTGATAAGTTTGAAGTGCGTGTGAAAAGTAAACAATTTAAA
 35 GATGGGGAGTGGCTGTACCATATAAGTCCTAAAACCTGGCTTCATTCTGTTTCGA
 TAGGCGGATCTAAGAACCCTTTTCATTGAAAAAGTTATCGCTAACGTATTTAGCTA
 CTTTAAGCCTAACATGGACGACTACTGCAATAGAACTTGTTTCGTTATTGATATT
 40 TTCTCTGATGATGCCTACCATTCTCAGGAGGACAGCGTTACCGAACATCGTGGCA
 ACAGAAGATTGAGTTTTTCATTGCGACAGAATTGAAGAAGTTCCCAAAACAGGGC
 TGGGCTCCTCGGCAGGTTTAGTCACAGTTTTAACTACAGCTTTGGCCTCCTTTTTT
 45 GTATCGGACCTGGAAAATAATGTAGACAAATATAGAGAAGTTATTCATAATTTAT
 CACAAGTTGCTCATTGTCAAGCTCAGGGTAAAATTGGAAGCGGGTTTGATGTAGC
 GGCGGCAGCATATGGATCTATCAGATATAGAAGATTCCCACCCGCATTAATCTCT
 AATTTGCCAGATATTGGAAGTGCTACTTACGGCAGTAAACTGGCGCATTTGGTTA
 50 ATGAAGAAGACTGGAATATAACGATTAAAAGTAACCATTTACCTTCGGGATTAA
 CTTTATGGATGGGCGATATTAAGAATGGTTCAGAAACAGTAAAACTGGTCCAGA

AGGTAAAAAATTGGTATGATTTCGCATATGCCGGAAGCTTGAAAATATATACAG
AACTCGATCATGCAAATTCTAGATTTATGGATGGACTATCTAAACTAGATCGCTT
5 ACACGAGACTCATGACGATTACAGCGATCAGATATTTGAGTCTCTTGAGAGGAAT
GACTGTACCTGTCAAAAGTATCCTGAGATCACAGAAGTTAGAGATGCAGTTGCC
ACAATTAGACGTTCCCTTTAGAAAAATAACTAAAGAATCTGGTGCCGATATCGAAC
10 CTCCCGTACAAACTAGCTTATTGGATGATTGCCAGACCTTAAAAGGAGTTCTTAC
TTGCTTAATACCTGGTGCTGGTGGTTATGACGCCATTGCAGTGATTGCTAAGCAA
GATGTTGATCTTAGGGCTCAAACCGCTGATGACAAAAGATTTTCTAAGGTTCAAT
GGCTGGATGTAACTCAGGCTGACTGGGGTGTTAGGAAAGAAAAAGATCCGGAAA
15 CTTATCTTGATAAATAACTTAAGGTAGATAATAGTGGTCCATGTGACATCTTTAT
AAATGTGAAGTTTGAAGTGACCGCGCTTAACATCTAACCATTTCATCTTCCGATAG
TACTTGAAATTGTTCCFTTCGGCGGCATGATAAAATTCTTTAATGGGTACAAGCT
20 ACCCGGGAAGATTCTCTTTTTTTATGATATTTGTACATAAACTTTATAAATGAAA
TTCATAATAGAAACGACACGAAATTACAAAATGGAATATGTTTCATAGGGTAGAC
GAAACTATATACGCAATCTACATACATTTATCAAGAAGGAGAAAAAGGAGGATG
TAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTCAACGTGATAAGGAAAA
25 AGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCACACAAAAAGT
TAGGTGTAACAGAAAATCATGAAACTATGATTCCCTAATTTATATATTGGAGGATT
TTCTCTAAAAAATAACAATAAAAAACACTCAATGACCTGACCAT
30 TTGATGGAGTTTAAGTCAATACCTTCTTGAACCATTTCCCATAAATGGTGAAAGTT
CCCTCAAGAATTTTACTCTGTCAGAAACGGCCTTAACGACGTAGTCGACCTCCTC
TTCAGTACTAAATCTACCAATACCAATCTGATGGAAGAATGGGCTAATGCATCA
35 TCCTTACCCAGCGCATGTAAACATAAGAAGGTTCTAGGGAAGCAGATGTACAG
GCTGAACCCGAGGATAATGCGATATCCCTTAGTGCCATCAATAAAGATTCTCCTT
CCACGTAGGCGAAAGAAACGTTAACACGTTTAAAC

SEQ ID NO: 52

pAM493 sequence (excluding vector backbone)

GTTTAAACTACTCAGTATATTAAGTTTCGAATTGAAGGGCGAACTCTTATTCGAA
45 GTCGGAGTCACCACAACACTTCCGCCATACTCTCCGAATCCTCGTTTCCTAAAG
TAAGTTTACTTCCACTTGTAGGCCTATTATTAATGATATCTGAATAATCCTCTATT
AGGGTTGGATCATTCAGTAGCGCGTGCGATTGAAAGGAGTCCATGCCCGACGTC
50 GACGTGATTAGCGAAGGCGCGTAACCATTGTCATGTCTAGCAGCTATAGAATA
ACCTCCTTGACACCACTTGCGGAAGTCTCATCAACATGCTCTTCCTTATTACTCAT

TCTCTTACCAAGCAGAGAATGTTATCTAAAACTACGTGTATTTACCTCTTTCTC
 GACTTGAACACGTCCAACCTCTTAAGTACTACCACAGCCAGGAAAGAATGGATC
 5 CAGTTCTACACGATAGCAAAGCAGAAAACACAACCAGCGTACCCCTGTAGAAGC
 TTCTTTGTTTACAGCACTTGATCCATGTAGCCATACTCGAAATTTCAACTCATCTG
 AAACCTTTTCCTGAAGGTTGAAAAAGAATGCCATAAGGGTCACCCGAAGCTTATTC
 10 ACGCCCGGGAGTTATGACAATTACAACAACAGAATTCTTTCTATATATGCACGAA
 CTTGTAATATGGAAGAAATTATGACGTACAACTATAAAGTAAATATTTTACGTA
 ACACATGGTGCTGTTGTGCTTCTTTTTCAAGAGAATACCAATGACGTATGACTAA
 GTTTAGGATTTAATGCAGGTGACGGACCCATCTTTCAAACGATTTATATCAGTGG
 15 CGTCCAAATTGTTAGGTTTTGTTGGTTCAGCAGGTTTCCTGTTGTGGGTCATATGA
 CTTTGAACCAAATGGCCGGCTGCTAGGGCAGCACATAAGGATAATTCACCTGCC
 AAGACGGCACAGGCAACTATTCTTGCTAATTGACGTGCGTTGGTACCAGGAGCG
 20 GTAGCATGTGGGCCTCTTACACCTAATAAGTCCAACATGGCACCTTGTGGTTCTA
 GAACAGTACCACCACCGATGGTACCTACTTCGATGGATGGCATGGATACGGA
 TTCTCAAATCACCGTCCACTTCTTTCATCAATGTTATACAGTTGGAACTTTCGACA
 TTTTGTGCAGGATCTTGTCTAATGCCAAGAAAACAGCTGTCACTAAATTAGCTG
 25 CATGTGCGTTAAATCCACCAACAGACCCAGCCATTGCAGATCCAACCAAAT
 TCTTAGCAATGTTCAACTCAACCAATGCGGAAACATCACTTTTTAACACTTTTCTG
 ACAACATCACCAGGAATAGTAGCTTCTGCGACGACACTCTTACCACGACCTTCGA
 30 TCCAGTTGATGGCAGCTGGTTTTTTGTGCGGTACAGTAGTTACCAGAAACGGAGAC
 AACCTCCATATCTTCCCAGCCATACTCTTCTACCA'TTTGCTTTAATGAGTATTCGA
 CACCCTTAGAAATCATATTCATACCCATTGCGTCACCAGTAGTTGTTCTAAATCTC
 35 ATGAAGAGTAAATCTCCTGCTAGACAAGTTTGAATATGTTGCAGACGTGCAAATC
 TTGATGTAGAGTTAAAAGCTTTTTTAATTGCGTTTTGTCCCTCTTCTGAGTCTAAC
 CATATCTTACAGGCACCAGATCTTTTCAAAGTTGGGAAACGGACTACTGGGCCTC
 40 TTGTCATACCATCCTTAGTTAAACAGTTGTTGCACCACCGCCAGCATTGATTGC
 CTTACAGCCACGCATGGCAGAAGCTACCAAACAACCCTCTGTAGTTGCCATTGGT
 ATATGATAAGATGTACCATCGATAACCAAGGGGCCTATAACACCAACGGGCAAA
 45 GGCATGTAACCTATAACATTTTCACAACAAGCGCCAAATACGCGGTCGTAGTCAT
 AATTTTTATATGGTAAACGATCAGATGCTAATACAGGAGCTTCTGCCAAAATTGA
 AAGAGCCTTCCTACGTACCGCAACCGCTCTCGTAGTATCACCTAATTTTTTCTCCA
 50 AAGCGTACAAAGGTAACCTACCGTGAATAACCAAGGCAGCGACCTCTTTGTTCTT
 CAATTGTTTTGTATTTCCACTACTTAATAATGCTTCTAATTCTTCTAAAGGACGTA
 TTTTCTTATCCAAGCTTTCAATATCGCGGGAATCATCTTCCTCACTAGATGATGAA

GGCCTGATGAGCTCGATTGCGCAGATGATAAACTTTTGACTTTTCGATCCAGAAA
 TGACTGTTTTATTGGTTAAAACTGGTGTAGAAGCCTTTTGTACAGGAGCAGTAAA
 5 AGACTTCTTGGTGACTTCAGTCTTCACCAATTGGTCTGCAGCCATTATAGTTTTTT
 CTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTTTATG
 ACATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAAAGT
 10 GGTTATGCAGCTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCTTCG
 CTGATTAATTACCCAGAAAATAAGGCTAAAAAACTAATCGCATTATTATCCTATG
 GTTGTTAATTTGATTCGTTGATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAATTT
 15 TTCCTCTTCATAACCATAAAAAGCTAGTATTGTAGAATCTTTATTGTTCCGGAGCAGT
 GCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACCAGGACGC
 ACGGAGGAGAGTCTTCCGTCGGAGGGCTGTCGCCCGCTCGGCGGCTTCTAATCCG
 TACTTCAATATAGCAATGAGCAGTTAAGCGTATTACTGAAAGTTCCAAAGAGAA
 20 GGTTTTTTTAGGCTAAGATAATGGGGCTCTTTACATTTCCACAACATATAAGTAA
 GATTAGATATGGATATGTATATGGTGGTATTGCCATGTAATATGATTATTAACT
 TCTTTGCGTCCATCCAAAAAAAAGTAAGAATTTTTGAAAATTCAATATAAATGA
 25 CTGCCGACAACAATAGTATGCCCCATGGTGCAGTATCTAGTTACGCCAAATTAGT
 GCAAAACCAAACACCTGAAGACATTTTGGAAGAGTTTCCTGAAATTATTCCATTA
 CAACAAAGACCTAATACCCGATCTAGTGAGACGTCAAATGACGAAAGCGGAGAA
 30 ACATGTTTTTCTGGTCATGATGAGGAGCAAATTAAGTTAATGAATGAAAATTGTA
 TTGTTTTGGATTGGGACGATAATGCTATTGGTGCCGGTACCAAGAAAGTTTGTCA
 TTTAATGGAAAATATTGAAAAGGGTTTACTACATCGTGCATTCTCCGTCTTTATTT
 35 TCAATGAACAAGGTGAATTACTTTTACAACAAAGAGCCACTGAAAAAATAACTT
 TCCCTGATCTTTGGACTAACACATGCTGCTCTCATCCACTATGTATTGATGACGAA
 TTAGGTTTGAAGGGTAAGCTAGACGATAAGATTAAGGGCGCTATTACTGCGGCG
 GTGAGAAAACCTAGATCATGAATTAGGTATTCCAGAAGATGAAACTAAGACAAGG
 40 GGTAAGTTTCACTTTTTAAACAGAATCCATTACATGGCACCAAGCAATGAACCAT
 GGGGTGAACATGAAATTGATTACATCCTATTTTATAAGATCAACGCTAAAGAAA
 ACTTGACTGTCAACCCAAACGTCAATGAAGTTAGAGACTTCAAATGGGTTTCACC
 45 AAATGATTTGAAAACCTATGTTTGCTGACCCAAGTTACAAGTTTACGCCTTGGTTT
 AAGATTATTTGCGAGAATTACTTATTCAACTGGTGGGAGCAATTAGATGACCTTT
 CTGAAGTGGAATAATGACAGGCAAATTCATAGAATGCTATAACAACGCGTCAATA
 50 ATATAGGCTACATAAAAATCATAATAACTTTGTTATCATAGCAAAATGTGATATA
 AAACGTTTCATTTACCTGAAAAATAGTAAAAATAGGCGACAAAAATCCTTAGT
 AATATGTAACTTTATTTTCTTTATTTACCCGGGAGTCAGTCTGACTCTTGCGAGA

GATGAGGATGTAATAATACTAATCTCGAAGATGCCATCTAATACATATAGACATA
 CATATATATATATATACATTCTATATATTCTTACCCAGATTCTTTGAGGTAAGACG
 5 GTTGGGTTTTATCTTTTGCAGTTGGTACTATTAAGAACAATCGAATCATAAGCATT
 GCTTACAAAGAATACACATACGAAATATTAACGATAATGTCAATTACGAAGACT
 GAACTGGACGGTATATTGCCATTGGTGGCCAGAGGTAAAGTTAGAGACATATAT
 10 GAGGTAGACGCTGGTACGTTGCTGTTTGTGCTACGGATCGTATCTCTGCATATG
 ACGTTATTATGGAAAACAGCATTCTGAAAAGGGGATCCTATTGACCAAACCTGTC
 AGAGTTCTGGTTCAAGTTCCTGTCCAACGATGTTCTGTAATCATTGTTGGTTCGACATC
 15 GCCCCAGGTAAGACTATTTTCGATTATCTACCTGCAAAATTGAGCGAACCAGT
 ACAAACGCAACTAGAAGACCGCTCTCTATTGGTTCACAAACATAAACTAATTCC
 ATTGGAAGTAATTGTCAGAGGCTACATCACCGGATCTGCTTGGAAGAGTACGT
 20 AAAAACAGGTACTGTGCATGGTTTGAAACAACCTCAAGGACTTAAAGAATCTCA
 AGAGTTCCCAGAACCAATCTTCACCCCATCGACCAAGGCTGAACAAGGTGAACA
 TGACGAAAACATCTCTCCTGCCAGGCCGCTGAGCTGGTGGGTGAAGATTTGTCA
 CGTAGAGTGGCAGAACTGGCTGTAAAACCTGTACTCCAAGTGCAAAGATTATGCT
 25 AAGGAGAAGGGCATCATCATCGCAGACACTAAATTGTTTAAAC

SEQ ID NO: 53

pAM426 sequence

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC
 GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC
 35 GTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCA
 GATTGTACTGAGAGTGCACCATATCGACTACGTCGTAAGGCCGTTTCTGACAGAG
 TAAAATTCTTGAGGGAACCTTTCACCATATGGGAAATGCTTCAAGAAGGTATTGA
 40 CTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTTGTATTTTTT
 TTTTTTAGAGAAAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTTTCT
 GTTACACCTAACTTTTTGTGTGGTGGCCTCCTCCTTGTCATATTAATGTTAAAGT
 GCAATTCCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATT
 45 CCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAG
 TTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTTCGTGTCGTTTCTATTATGA
 ATTTCAATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTAAAG
 50 CAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTT
 GGAACCACCTAAATCACCAGTTCTGATACCTGCATCCAAAACCTTTTTAACTGCA
 TCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCATTGC

AGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGC
AGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGA
5 GGCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTT
CATCGGAGATGATATCACCAACATGTTGCTGGTGATTATAATACCATTTAGGTG
GGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACC
10 TTCAATGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCT
TGAAGAGGCCAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTC
ATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTGATTCTTTGCACTTCTGGAACG
GTGTATTGTTCACTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACC
15 AAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCA
AATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACAT
GGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTT
20 AGGTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAAC
GGCATCAACCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTA
GCATCGATAGCAGCACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGG
25 AACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTG
ACCAACGTGGTCACCTGGCAAACGACGATCTTCTTAGGGGCAGACATTACAAT
GGTATATCCTTGAAATATATATAAAAAAAGGCGCCTTAGACCGCTCGGCCAAAC
AACCAATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTT
30 GAACACACATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGAT
TTTGATGTAATTGTTGGGATTCCATTTTTAATAAGGCAATAATATTAGGTATGTGG
ATATACTAGAAGTTCTCCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGAT
35 GCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAACGTTAATATTTTGTTAAA
ATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTTAACCAATAGGCCGAAATCG
GCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTT
40 CAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGC
GAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGGAAGATCCGAGGCCTAGC
TTTAACGAACGCAGAATTTTCGAGTTATTAACTTAAATACGCTGAACCCGAAC
ATAGAAATATCGAATGGGAAAAAAAAAACTGCATAAAGGCATTAAAAGAGGAGC
45 GAATTTTTTTTTTAATAAAAATCTTAATAATCATTAAAAGATAAATAATAGTCTAT
ATATACGTATATAAATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTAGC
GTAAAGGATGGGGAAAGAGAAAAAGAAAAAAATTGATCTATCGATTTCAATTCAA
50 TTCAATTTATTTCTTTTCGGATAAGAAAGCAACACCTGGCAATTCCTTACCTTCCA
ATAATTCCAAAGAAGCACCACCACCAGTAGAGACATGGGAGACCCGGGCCATGG

TTAGATAGACATAGGGTAAACTAGCAATGATTTGATCAAATGCTTGTATTCATCT
 CCCATTCTCGTAAAATTGTCTTTACCTGCATATTGGACCTCTAAAAATTGGCAAA
 5 GATATATAACAGCCATAAGTAAAGGTCTTGGGATATTCTTTGTTGTTAAATACTC
 TCTGTTTATGTCTTTCCAAACGTCCCTCCACTTCCTTATAAATCAGTGTCTGAGCAT
 ATTCTTCGTTGACATTGTATTCCCTTCATGTAAGATTCTAAAGAGCTTGAACATATGT
 10 TTTCTCTCCTGTTCCGCTTTATGAGTCATCAGGTCATTTAATCTCCTACCCAGAAT
 ACCACTGTAACGGAATAAAGGCGGAGCAGATACAGCCCACTCAACTGATTCCTT
 AGTGAAAATATCGCTCATTCCCTAGATAACAGGTAGTTGTTAGCAAGTTTGCACCA
 15 CCAGTGATAATAACTACGGGATCGTGCTCTTCAGTTGTCGGTATGTGTCCTTCATT
 AGCCCATTTTCGCTTCTACCATTAGATTCCCTTACGAATTCTTTAACGAACTCCTTCC
 CACAGTTGAATAAATCAGTTCTACCTTCTTTGGCCAGAAACTCCTCCATTTCTGTG
 TAGGTATCCATGAATAATTTGTAAATAGGCTTCATGTATTCCGGCAACGTGTCTA
 20 AGCAGGTGATCGACCATCTTTCCACGGCTTCAGTGAAAATCTTTAACTCCTCGTA
 AGTTCCATATGCGTCATACGTGTCATCAATAAGTGTTATCACAGCAACTGCCTTA
 GTGAAAAAAACTCTAGCTCTTGAATACTGGGGTTCGTAACCAGAACCTAAACCCC
 25 AAAAATAGCATTCAACGATACGATCTCTCAGACATGGGGCATTTTTCTTAATATC
 AAATGCCTTCCACCACTTGCATACGTGACTCAACTCTTCCTTATGTAGGCTCTGCA
 ATAGATTGAACTCCAGTTTAGCTAACTTTAGCAGAGTTTTATTATGGGAGTCTTGT
 30 TGCTGATAGAAGGGTATGTACTGGGCGGCCTCGATCCTTGGCAATCTCTTCCACA
 ATGGTTGCTTTAAAGCTCTCTGGATTTCAGTGAATAAAGCGGGGTTTGTACTAAA
 CGCGTCCTTTGTGCATAATCGATAGCCTTGATCTTGTGAATCCCAGGGCATCTTCAA
 GAATTATTTGCCCCGGAACCTCTCATGGACGTAGCCTCATATAATCCAACAATCC
 35 TTCAACATCATTGCTAACGATTGTTTAAAAGCACCATTCTTGTCTTTATAGTTAT
 TAAACACATCACACGTGACATAGTATCCTTGTTTACGCATCAGCCTAAACCATAA
 GCTAGACCTGTGCGCCATTCCAATTATCACCATAGGTCTCGTAAATACATTGCAAT
 40 GCATGATCAATTTACGTTCAAATGATACGGAATACCTAAACGTTGAATCTCGT
 CAATCAGCTTCAACAAATTTGCATGTTTCATAGGAATATCCAATGCTTCCTTTAAC
 AACTGTCTTACTTCCTTCTTTAGATCGTTTACTATTTGCTCCACACCCTGTTCAACT
 45 TGTTTCTCATAAATCAAAAATTGATCGCCCCAAATAGAAGGTGGGAAATTTGCAA
 TTGGCCTTATAGGTTTCTCTTCAGTCAAGGCCATTGTTTTCTGCAGATCCGGGGTT
 TTTTCTCCTTGACGTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTT
 50 TATTACATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAA
 AGTGGTTATGCAGTTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCT
 TCGCTGATTAATTACCCAGAAATAAGGCTAAAAAACTAATCGCATTATCATCCT

ATGGTTGTTAATTTGATTCGTTCAATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAA
TTTTTCCTCTTCATAACCATAAAAAGCTAGTATTGTAGAATCTTTATTGTTTCGGAGC
5 AGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACGAGGA
CGCACGGAGGAGAGTCTTCCTTCGGAGGGCTGTCACCCGCTCGGCGGCTTCTAAT
CCGTACTAAGATCTGCTTTAATTTGGCCGGCGAACGTGGCGAGAAAGGAAGGGA
10 AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTG TAGCGGTCACGCTG
CGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGCTCGCGC
CATTCGCCATTACAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTT
15 CGCTATTACGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCGT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACA
GAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
20 CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCT
GACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
25 CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG
CGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG TAGGTCGTTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTT CAGCCCGACCGCTGCGCCTTATCCG
GTA ACTATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGC
30 AGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTT
CTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGC
GCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCA
35 AACA AACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
CAGTGGAACGAAAACCTCACGTAAAGGGATTTTGGTCATGAGATTATCAAAAAGG
40 ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTTAAATCAATCTAAAGTA
TATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT
CTCAGCGATCTGTCTATTTTCGTT CATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCG
45 AGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAG
GGCCGAGCGCAGAAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
TGTTGCCGGGAAGCTAGAGTAAGTAGTTGCGCAGTTAATAGTTTGCGCAACGTTG
50 TTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA

AAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGT
 GTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
 5 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG
 TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCA
 CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
 10 CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGAC
 CCAACTGATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAAC
 AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
 15 TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTC
 ATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCG
 CGCACATTTCCCGAAAAGTGCCACCTGAACGAAGCATCTGTGCTTCATTTTGTA
 20 GAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAACAAAGAATCTGAGCTGC
 ATTTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTACCAACGAAGAATC
 TGTGCTTCATTTTTGTAAAACAAAAATGCAACGCGAGAGCGCTAATTTTTCAAAC
 AAAGAATCTGAGCTGCATTTTTACAGAACAGAAATGCAACGCGAGAGCGCTATT
 25 TTACCAACAAAGAATCTATACTTCTTTTTTTGTTCTACAAAAATGCATCCCGAGAG
 CGCTATTTTTCTAACAAAGCATCTTAGATTACTTTTTTTCTCCTTTGTGCGCTCTAT
 AATGCAGTCTCTTGATAACTTTTTGCACTGTAGGTCCGTAAAGGTTAGAAGAAGG
 30 CTACTTTGGTGTCTATTTTTCTCTTCCATAAAAAAAGCCTGACTCCACTTCCCGCGT
 TTA CTGATTACTAGCGAAGCTGCGGGTGCATTTTTTCAAGATAAAGGCATCCCCG
 ATTATATTCTATACCGATGTGGATTGCGCATACTTTGTGAACAGAAAGTGATAGC
 35 GTTGATGATTCTTCATTGGTCAGAAAATTATGAACGGTTTCTTCTATTTTGTCTCT
 ATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTTCGATTCACTCTATG
 AATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAGAGATAAACATAA
 AAAATGTAGAGGTGAGTTTAGATGCAAGTTCAAGGAGCGAAAGGTGGATGGGT
 40 AGGTTATATAGGGATATAGCACAGAGATATATAGCAAAGAGATACTTTTGAGCA
 ATGTTTGTGGAAGCGGTATTCGCAATATTTTAGTAGCTCGTTACAGTCCGGTGCG
 TTTTTGGTTTTTTGAAAGTGCGTCTTCAGAGCGCTTTTGGTTTTCAAAGCGCTCT
 45 GAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGAACTTCAAAGCGT
 TTCCGAAAACGAGCGCTTCCGAAAATGCAACGCGAGCTGCGCACATACAGCTCA
 CTGTTACGTCGCACCTATATCTGCGTGTTGCCTGTATATATATATACATGAGAA
 50 GAACGGCATAGTGCGTGTTTATGCTTAAATGCGTACTTATATGCGTCTATTTATGT
 AGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTCATGCGGGGTA
 TCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTGCCACTCCTCAA

TTGGATTAGTCTCATCCTTCAATGCTATCATTTCTTTGATATTGGATCATACTAA
GAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCT
TTCGTC

SEQ ID NO: 54
pAM322 sequence

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC
GGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC
GTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACCTATGCGGCATCAGAGCA
GATTGTACTGAGAGTGCACCATATCGACTACGTCGTAAGGCCGTTTCTGACAGAG
TAAAATTCTTGAGGGAACTTTCACCATTATGGGAAATGCTTCAAGAAGGTATTGA
CTTAAACTCCATCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTTGTATTTTTT
TTTTTTAGAGAAAAATCCTCCAATATCAAATTAGGAATCGTAGTTTCATGATTTTCT
GTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCATATTAATGTTAAAGT
GCAATTCTTTTTCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTAAT
CCTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAG
TTTCGTCTACCCTATGAACATATTCCATTTTGTAATTTCTGTGTCGTTTCTATTATGA
ATTTCAATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTTTTTAAG
CAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTT
GGAACCACCTAAATCACCAAGTTCTGATACCTGCATCCAAAACCTTTTAACTGCA
TCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAATGACAATTTCAACATCATTGC
AGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGC
AGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGA
GGCCAAGGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTT
CATCGGAGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTG
GGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACC
TTCAATGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCT
TGAAGAGGCCAAAAGATTAGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTC
ATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTTGATTCTTTGCACTTCTGGAACG
GTGTATTGTTCACTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACC
AAAGTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCA
AATTGTGGCTTGATTGGAGATAAGTCTAAAAGAGAGTCGGATGCAAAGTTACAT
GGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTT
AGGTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAAC

GGCATCAACCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTA
GCATCGATAGCAGCACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGG
5 AACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTG
ACCAACGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGACATTACAAT
GGTATATCCTTGAAATATATATAAAAAAAGGCGCCTTAGACCGCTCGGCCAAAC
10 AACCAATTACTTGTTGAGAAATAGAGTATAATTATCCTATAAATATAACGTTTTT
GAACACACATGAACAAGGAAGTACAGGACAATTGATTTTGAAGAGAATGTGGAT
TTTGATGTAATTGTTGGGATTCCATTTTAAATAAGGCAATAATATTAGGTATGTGG
ATATACTAGAAGTTCTCCTCGACCGTCGATATGCGGTGTGAAATACCGCACAGAT
15 GCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAACGTTAATATTTTGTAAA
ATTCGCGTTAAATTTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCG
GCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTT
20 CAGTTTGGACAAGAGTCCACTATTAAGAAGCGTGGACTCCAACGTCAAAGGGC
GAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGGAAGATCCGAGGCCTAGC
TTTAACGAACGCAGAATTTTCGAGTTATTAACCTTAAAATACGCTGAACCCGAAC
25 ATAGAAATATCGAATGGGAAAAAAACTGCATAAAGGCATTAAAAGAGGAGC
GAATTTTTTTTTAATAAAAAATCTTAATAATCATTAAAAGATAAATAATAGTCTAT
ATATACGTATATAAATAAAAAATATTCAAAAAATAAAATAAACTATTATTTTAGC
30 GTAAAGGATGGGGAAAGAGAAAAAGAAAAAAATTGATCTATCGATTTCATTCAA
TTCAATTTATTTCTTTTCGGATAAGAAAGCAACACCTGGCAATTCCTTACCTTCCA
ATAATTCCAAAGAAGCACCACCACCAGTAGAGACATGGGAGACCCGGGCCATGG
TTAGATAGACATAGGGTAAACTAGCAATGATTTGATCAAATGCTTGTATTTCATCT
35 CCCATTCTCGTAAAATTGTCTTTACCTGCATATTGGACCTCTAAAATTTGGCAAA
GATATATAACAGCCATAAGTAAAGGTCTTGGGATATTCTTTGTTGTAAATACTC
TCTGTTTATGTCTTTCCAAACGTCCTCCACTTCCTTATAAATCAGTGTCTGAGCAT
40 ATTCTTCGTTGACATTGTATTCCTTCATGTAAGATTCTAAAGAGCTTGAACATATGT
TTTCTCTCCTGTTCCGCTTTATGAGTCATCAGGTCATTTAATCTCCTACCCAGAAT
ACCACTGTAACGGAATAAAGGCGGAGCAGATACAGCCCACTCAACTGATTCCTT
45 AGTGAAAATATCGCTCATTCCTAGATAACAGGTAGTTGTTAGCAAGTTTGCACCA
CCAGTGATAATAACTACGGGATCGTGCTCTTCAGTTGTTCGGTATGTGTCCTTCATT
AGCCCATTTTCGCTTCTACCATTAGATTCCTTACGAATTCTTTAACGAACCTCCTTCC
50 CACAGTTGAATAAATCAGTTCTACCTTCTTTGGCCAGAACTCCTCCATTTCTGTG
TAGGTATCCATGAATAATTTGTAAATAGGCTTCATGTATTCCGGCAACGTGTCTA
AGCAGGTGATCGACCATCTTTCCACGGCTTCAGTGAAAATCTTTAACTCCTCGTA
55

AGTTCCATATGCGTCATACGTGTCATCAATAAGTGTTATCACAGCAACTGCCTTA
 GTGAAAAAAACTCTAGCTCTTGAATACTGGGGTTCGTAACCAGAACCTAAACCCC
 5 AAAAAATAGCATTCAACGATACGATCTCTCAGACATGGGGCATTCTTTCTTAATATC
 AAATGCCTTCCACCACCTTGCATACGTGACTCAACTCTTCCTTATGTAGGCTCTGCA
 ATAGATTGAACTCCAGTTTAGCTAACTTTAGCAGAGTTTTATTATGGGAGTCTTGT
 10 TGCTGATAGAAGGGTATGTACTGGGCGGCCTCGATCCTTGGCAATCTCTTCCACA
 ATGGTTGCTTTAAAGCTCTCTGGATTTCAGTGAATAAAGCGGGGTTTGTACTAAA
 CGCGTCCTTTGTCATAATCGATAGCCTTGATCTTGTGAATCCCAGGGGCATCTTCAA
 GAATTATTTGCCCCGGAACCTCTCATGGACGTAGCCTCATATAATTCCAACAATCC
 15 TTCAACATCATTGCTAACGATTGTTTAAAAGCACCATTCTTGTCTTTATAGTTAT
 TAAACACATCACACGTGACATAGTATCCTTGTTTACGCATCAGCCTAAACCATAA
 GCTAGACCTGTCGCCATTCCAATTATCACCATAGGTCTCGTAAATACATTGCAAT
 20 GCATGATCAATTTACGTTCAAAATGATACGGAATACCTAAACGTTGAATCTCGT
 CAATCAGCTTCAACAAATTTGCATGTTTCATAGGAATATCCAATGCTTCCTTTAAC
 AACTGTCTTACTTCCTTCTTTAGATCGTTTACTATTTGCTCCACACCCTGTTCAACT
 25 TGTTTCTCATAAATCAAAAATTGATCGCCCCAAATAGAAGGTGGGAAATTTGCAA
 TTGGCCTTATAGGTTTCTCTTCAGTCAAGGCCATTGTTTTCTGCAGATCCGGGGTT
 TTTTCTCCTTGACGTTAAAGTATAGAGGTATATTAACAATTTTTTGTGATACTTT
 30 TATTACATTTGAATAAGAAGTAATACAAACCGAAAATGTTGAAAGTATTAGTTAA
 AGTGGTTATGCAGTTTTTGCATTTATATATCTGTTAATAGATCAAAAATCATCGCT
 TCGCTGATTAATTACCCCGAGAAATAAGGCTAAAAAACTAATCGCATTATCATCCT
 35 ATGGTTGTTAATTTGATTTCGTTCAATTTGAAGGTTTGTGGGGCCAGGTTACTGCCAA
 TTTTCTCCTTTCATAACCATAAAAGCTAGTATTGTAGAATCTTTATTGTTCCGAGC
 AGTGCGGCGCGAGGCACATCTGCGTTTCAGGAACGCGACCGGTGAAGACGAGGA
 CGCACGGAGGAGAGTCTTCCTTCGGAGGGCTGTCACCCGCTCGGCGGCTTCTAAT
 40 CCGTACTAAGATCTGCTTTAATTTGGCCGGCGAACGTGGCGAGAAAGGAAGGGA
 AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTG
 CGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCGCGC
 45 CATTCGCCATTCAAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTT
 CGCTATTACGCCAGCTGAATTGGAGCGACCTCATGCTATACCTGAGAAAGCAACC
 TGACCTACAGGAAAGAGTTACTCAAGAATAAGAATTTTCGTTTTAAACCTAAGA
 50 GTCACTTTAAAATTTGTATACACTTATTTTTTTTATAACTTATTTAATAATAAAAA
 TCATAAATCATAAGAAATTCGCTTATTTAGAAAGTGTCAACAACGTATCTACCAAC
 GATTTGACCCTTTTCCATCTTTTCGTAAATTTCTGGCAAGGTAGACAAGCCGACA
 55

ACCTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAATTGTTAATTAAGAGTCA
 GTCGACTTAAAACTAGGGACCAATAGCAATTCTGTTTTACGTTGCATTGTTGCA
 5 CCTGAACTTTCCGTCATGTCAATTTGATCATATGAAACTCCATTGGGCAACTTCCA
 GTTGAAATGATAAAGAATGTTGGCTAGTGGCAGTTGAACATTGGCCAAACCTAA
 CGCAGCGCCAGGACACATACGACGTCCAGCCCCAAATGGTAAATATTCATATTC
 10 GGCGCCCATCACTGTTGCCGAAGAGTTTTCAAATCTTTCAGGTATAAACGCTTCT
 GCATCCTTCCAGTATTCAGGATCTCTATTGATCGCAAACACATTAACGATTAATTT
 CGTTTTGTTAGGGATATTATAACCAGCCAAGTTTACTGGCTGACGACATTCTCTA
 15 GGTAGCACTAACGGCAAGGGTGGGTGTAGTCTAAGAGTCTCTTTGATGACCATAT
 TCAAGTAGGACAATTCTTGTATATCTTCTTCATGTATTTTTCTTTCCCATTC AAG
 GCCTTACGTAATTCAGCCTGAACCTTTTCCATTGCTTTTCGGACATTTTATTAGCTC
 GCTTATAGCCCATTCTATGGTAGAACTTGAAGTGTTCGGTCCCTGCACCGAACATG
 20 TCCAAAATTATTGCTTTGATATTATCCGAAGTCAGAGGAAACTCAGCAGAATCCT
 TTAATCTAAGTAATACATCTAATAGGGTTTTCGTTGGTTTTGGATGACGTATTTACG
 GTATGTTTACGTACCAAATTGTCAATTAAGTTATCAATCTTTTTACGTAGGCTAGT
 25 TAATCTTGCTCTCTTACCGCTCAAGTGATGCAAGAACTTTTTAGATGGGAAAATA
 TCGGCAACATCGAAACCGCCTGTTTGTCTCAGTATTTCTTTAACAATTTTCAAGTAAG
 TTCCTTTTGATCTTTAATTCCTTACCAAACGCAGCACGGGATAGTATAGTGGCA
 30 ATTAGTTTAAAAACGTTTTCACTTAAATTTACTGGTCTACCACTACCTGAAGCCTT
 TATTTCTGGACTAAATTCCAACATTCTTCTTCCCTCAACGATTGAAATGACTTAA
 CCTTTTTTACAGACAACAATTCAAGAGTACAAATCTTCTTAATTGTCTCCAGTAT
 35 TCCCCATATGGAGCAAGGACAACATCAGTGTTATGATATAAAACTATTTCCCCAG
 TTAAAGTTTTCGGGTCTATTAGCGAAAGTAATATCGTAGGTTGTAAGAATTTCTT
 AGCCCACTTAGGACTCGACACGACTATTGTGGGTACCTCTCCCAATTGAAGGTGC
 ATTAGCGAACCATATTTTCTCGCTAAATCCCTTACACCCCTGTGTGGTGTGGTTCC
 40 GATCAAATGGTGCATGTGACCAATGATGGGTAGCCTCCAAGGTTCCGGCAAGGA
 CTTTTTAGTTGACTTACTTCTAGTGGCAAATTTGTACACGAACAACAAAATAGTT
 GCTAAAGCAATTGATGTAGTTAAAGATAGTGCCATAGCCTTTAAAATTGACTTCA
 45 TTGTTTTCTAGGCCTTTAGTGAGGGTTGAATTCGAATTTTCAAAAATTCTTACTT
 TTTTTTTGGATGGACGCAAAGAAGTTTAATAATCATATTACATGGCATTACCACC
 ATATACATATCCATATACATATCCATATCTAATCTTACTTATATGTTGTGGAAATG
 50 TAAAGAGCCCCATTATCTTAGCCTAAAAAAACCTTCTCTTTGGAACCTTTCAGTAA
 TACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCCGCCGAGCGGG
 TGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTTCG

GTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCT
 ACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCA
 5 CAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTA
 GTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTA
 TTAACAGATATATAAATGCAAAAACCTGCATAACCACTTTAACTAATACTTTCAAC
 10 ATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAATT
 GTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACCCCAAGCTTCCCGGG
 AAAACAATGCAATCGACAACCTTCCGTAAACTATCACCTTTCGATCTTATGACTG
 15 CCTTGTTAAATGGTAAAGTTAGTTTCGACACGTCCAATACTTCCGATACAAATAT
 ACCACTGGCGGTTTTTCATGGAAAACAGGGAATTGCTTATGATATTAACAACCAGT
 GTGGCCGTTTTAATTGGTTGTGTGGTTGTATTGGTATGGAGAAGATCATCAAGTG
 20 CCGCTAAGAAGGCCGCCGAATCACCAGTCATTGTCGTCCCAAAGAAAGTCACTG
 AAGATGAGGTTGATGACGGCAGAAAGAAAGTTACTGTATTTTTTCGGGACACAAA
 CGGGGACTGCGGAAGGTTTTGCGAAAGCTCTAGTTGAAGAAGCCAAGGCAAGGT
 ACGAAAAAGCAGTATTCAAAGTTATTGATTTAGATGACTACGCCGCAGAAGATG
 25 ATGAATACGAAGAAAAGCTAAAGAAAGAATCTTTGGCATTCTTCTTTTAGCTAC
 CTATGGTGACGGAGAACCAACAGATAACGCCGCTAGATTCTATAAATGGTTTACT
 GAAGGAGAAGAAAAAGGTGAGTGGTTAGATAAGTTACAATACGCTGTCTTTGGA
 30 TTGGGAAATCGTCAATATGAACACTTCAATAAGATTGCAAAAGTGGTCGATGAA
 AAATTAGTTGAGCAGGGGGCTAAAAGGTTAGTGCCTGTCGGTATGGGTGATGAC
 GATCAATGTATCGAAGATGATTTTACTGCTTGAAGGAATTGGTTTGGCCAGAAT
 35 TAGATCAGCTATTGAGGGACGAAGATGACACAAGTGTGCTACTCCGTACACCG
 CCGCTGTTGGCGAATATCGTGTTGTTTTTCACGATAAACCTGAAACTTACGATCA
 AGATCAATTGACCAACGGACACGCAGTTCACGACGCCCAACACCCATGCAGATC
 GAACGTTGCGGTCAAGAAAGAATTACACAGTCCCTTATCCGATAGGAGTTGTACT
 40 CATTTAGAATTTGATATTTCCAATACTGGACTATCGTATGAAACTGGCGACCATG
 TCGGTGTATATGTGGAAAACCTGTCTGAAGTTGTAGATGAAGCCGAAAAATTGAT
 TGGGCTTCCTCCACATACATACTTTTCTGTGCATACAGATAATGAAGATGGTACT
 45 CCACTTGGCGGAGCCTCGTTACCACCTCCCTTTCCACCATGTACACTTAGAAAAG
 CTCTTGCACTCTTATGCAGATGTACTTTCTTCACCAAAGAAAAGTGCATTACTAGCT
 CTAGCCGCCCATGCTACCGACTCTACTGAAGCTGACCGTTTGAAATTCTTTGCTTC
 50 ACCTGCTGGCAAAGACGAGTACGCACAGTGGATTGTGGCATCTCACAGATCATT
 GCTGGAAGTGATGGAAGCCTTCCCATCGGCAAAGCCACCATTAGGCGTGTTTTTC
 GCATCTGTTGCCCCACGTTTACAGCCTAGATACTATTCCATATCTTCTAGCCCAA

ATTTGCCCCCAATCGTATTCATGTGACGTGTGCGCTGGTGTATGAACAAACTCCA
 TCAGGAAGGGTACATAAAGGTGTCTGTAGTACATGGATGAAAAACGCGGTGCCA
 5 ATGACTGAATCTCAAGATTGTTTCGTGGGCACCAATTTATGTTTCGTACTTCTAATTT
 TAGACTACCTAGTGACCCTAAAGTACCAGTGATTATGATCGGGCCTGGGACAGG
 ACTAGCGCCATTTCAGAGGTTTCTTACAAGAAAGATTGGCCCAAAAGGAAGCAGG
 10 TACGGAATTAGGAACCGCAATTCTATTCTTTGGTTGTCGTAATAGAAAAGTTGAC
 TTTATATACGAAGATGAGTTAAACAACCTTCGTTGAAACTGGAGCGTTATCAGAAT
 TAGTGACAGCATTCTCTAGGGAAGGTGCAACAAAAGAATACGTCCAACATAAAA
 15 TGACCCAAAAGGCCAGCGATATATGGAATTTGCTGTCCGAGGGTGCCTATTTGTA
 CGTTTGTGGTGATGCAAAGGGAATGGCTAAAGATGTTTACAGGACATTGCATAC
 AATTGTTTCAGGAACAAGGTTCTTGGATTCTCTAAGGCAGAACTTTATGTTAAA
 AACCTTCAGATGGCTGGTAGATATTTGCGTGATGTTTGGTGAGCTAGCTAAGATC
 20 CGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTAT
 TTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTT
 TTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGA
 25 AGGTTTTGGGACGCTCGAAGATCCAGCTGCATTAATGAATCGGCCAACGCGCGG
 GGAGAGGCGGTTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
 GCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
 30 ACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
 GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
 GGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC
 35 GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
 GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
 CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
 40 GGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
 TGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTAT
 CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCG
 GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAG
 45 TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAG
 CTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAG
 CAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
 50 CGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGA
 GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
 ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATC
 55

AGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACT
 CCCCCTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCT
 5 GCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAAC
 CAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCC
 ATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATA
 10 GTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTT
 GGTATGGCTTCATTACAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCC
 CCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAG
 15 TAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTT
 ACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGT
 CATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCGGCGTCAATACG
 GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACG
 20 TTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG
 TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTACCCAGCGTTTC
 TGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGA
 25 CACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTAT
 CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
 AAATAGGGGTTCGCGCACATTTCCCCGAAAAGTGCCACCTGAACGAAGCATCT
 30 GTGCTTCATTTTGTAGAACAAAAATGCAACGCGAGAGCGCTAATTTTCAAACAA
 AGAATCTGAGCTGCATTTTACAGAACAGAAATGCAACGCGAAAGCGCTATTTTA
 CCAACGAAGAATCTGTGCTTCATTTTGTAAAACAAAAAATGCAACGCGAGAGCG
 35 CTAATTTTCAAACAAAGAATCTGAGCTGCATTTTACAGAACAGAAATGCAACG
 CGAGAGCGCTATTTTACCAACAAAGAATCTATACTTCTTTTTTGTCTACAAAAAT
 GCATCCCGAGAGCGCTATTTTCTAACAAAGCATCTTAGATTACTTTTTTCTCCT
 TTGTGCGCTCTATAATGCAGTCTCTTGATAACTTTTTGCACTGTAGGTCCGTTAAG
 40 GTTAGAAGAAGGCTACTTTGGTGTCTATTTTCTCTTCCATAAAAAAAGCCTGACT
 CCACTTCCCGCGTTTACTGATTACTAGCGAAGCTGCGGGTGCAATTTTCAAGAT
 AAAGGCATCCCCGATTATATTCTATACCGATGTGGATTGCGCATACTTTGTGAAC
 45 AGAAAGTGATAGCGTTGATGATTCTTCATTGGTCAGAAAATTATGAACGGTTTCT
 TCTATTTTGTCTCTATATACTACGTATAGGAAATGTTTACATTTTCGTATTGTTTT
 GATTCACTCTATGAATAGTTCTTACTACAATTTTTTTGTCTAAAGAGTAATACTAG
 50 AGATAAACATAAAAAATGTAGAGGTCGAGTTTAGATGCAAGTTCAAGGAGCGAA
 AGGTGGATGGGTAGGTTATATAGGGATATAGCACAGAGATATATAGCAAAGAGA
 TACTTTTGAGCAATGTTTGTGGAAGCGGTATTCGCAATATTTTAGTAGCTCGTTAC
 55

AGTCCGGTGCGTTTTTTGGTTTTTTGAAAGTGCGTCTTCAGAGCGCTTTTGGTTTTTC
 AAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAGGA
 5 ACTTCAAAGCGTTTTCCGAAAACGAGCGCTTCCGAAAATGCAACGCGAGCTGCGC
 ACATACAGCTCACTGTTACGTCGCACCTATATCTGCGTGTTGCCTGTATATATAT
 ATACATGAGAAGAACGGCATAGTGCGTGTTTATGCTTAAATGCGTACTTATATGC
 10 GTCTATTTATGTAGGATGAAAGGTAGTCTAGTACCTCCTGTGATATTATCCCATTC
 CATGCGGGGTATCGTATGCTTCCTTCAGCACTACCCTTTAGCTGTTCTATATGCTG
 CCACTCCTCAATTGGATTAGTCTCATCCTTCAATGCTATCATTTCTTTGATATTG
 GATCATACTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTAT
 15 CACGAGGCCCTTTCGTC

Claims

1. A bio-organic compound production system comprising:
 - a. a vessel having a capacity of at least 100 liters;
 - b. an aqueous medium, within the vessel, forming a first phase;
 - 25 c. a plurality of genetically modified bacterial or fungal host cells, within the aqueous medium, capable of converting a carbohydrate carbon source present in the aqueous medium into at least one bio-organic compound having at least ten carbon atoms; and
 - d. a liquid organic second phase, comprising the at least one bio-organic compound in an amount of at least 90%, in contact with the first phase;
 - 30 wherein the at least one bio-organic compound is an isoprenoid compound; and
wherein the organic second phase is spontaneously formed.
2. The system as in claim 1, wherein the at least one bio-organic compound is a C₁₀ bio-organic compound, a C₁₅ bio-organic compound, a C₂₀ bio-organic compound, or a C₂₀+ bio-organic compound.
- 35 3. The system as in claim 1, wherein the vessel has a capacity of at least 1000 at least 10,000 liters, at least 50,000 liters, at least 100,000 liters, at least 500,000 liters, or at least 1,000,000 liters.

Patentansprüche

1. Herstellungssystem für eine bio-organische Verbindung, das Folgendes umfasst:
 - a. ein Gefäß mit einer Kapazität von zumindest 100 l;
 - 45 b. ein wässriges Medium innerhalb des Gefäßes, das eine erste Phase ausbildet;
 - c. eine Vielzahl genetisch modifizierter bakterieller oder pilzlicher Wirtszellen innerhalb des wässrigen Mediums, die in der Lage sind, eine im wässrigen Medium vorliegende Kohlenhydrat-Kohlenstoffquelle in zumindest eine bioorganische Verbindung, die zumindest zehn Kohlenstoffatome aufweist, umzuwandeln; und
 - 50 d. eine flüssige organische zweite Phase, die die zumindest eine bio-organische Verbindung in einer Menge von zumindest 90 % umfasst, in Kontakt mit der ersten Phase;
 - wobei die zumindest eine bio-organische Verbindung eine Isoprenoidverbindung ist; und
wobei die organische zweite Phase spontan ausgebildet wird.
2. System nach Anspruch 1, wobei die zumindest eine bio-organische Verbindung eine bio-organische C₁₀-Verbindung, eine bio-organische C₁₅-Verbindung, eine bio-organische C₂₀-Verbindung oder eine bio-organische C₂₀+ -Verbindung ist.
- 55 3. System nach Anspruch 1, wobei das Gefäß eine Kapazität von zumindest 1000 Litern, zumindest 10.000 Litern,

zumindest 50.000 Litern, zumindest 100.000 Litern, zumindest 500.000 Litern oder zumindest 1.000.000 Litern aufweist.

5 Revendications

1. Système de production de composé bioorganique comprenant :

- a. une cuve ayant une capacité d'au moins 100 litres;
 - b. un milieu aqueux, dans la cuve, formant une première phase ;
 - c. une pluralité de cellules hôtes bactériennes ou fongiques génétiquement modifiées, dans le milieu aqueux, capables de convertir une source de carbone glucidique présente dans le milieu aqueux en au moins un composé bioorganique ayant au moins dix atomes de carbone ; et
 - d. une deuxième phase organique liquide, comprenant l'au moins un composé bioorganique en une quantité d'au moins 90 %, en contact avec la première phase ;
- dans lequel l'au moins un composé bioorganique est un composé isoprénoïde ; et dans lequel la deuxième phase organique est formée spontanément.

2. Système selon la revendication 1, dans lequel l'au moins un composé bioorganique est un composé bioorganique en C₁₀, un composé bioorganique en C₁₅, un composé bioorganique en C₂₀ ou un composé bioorganique en C₂₀⁺.

3. Système selon la revendication 1, dans lequel la cuve a une capacité d'au moins 1000, au moins 10 000 litres, au moins 50 000 litres, au moins 100 000 litres, au moins 500 000 litres ou au moins 1 000 000 litres.

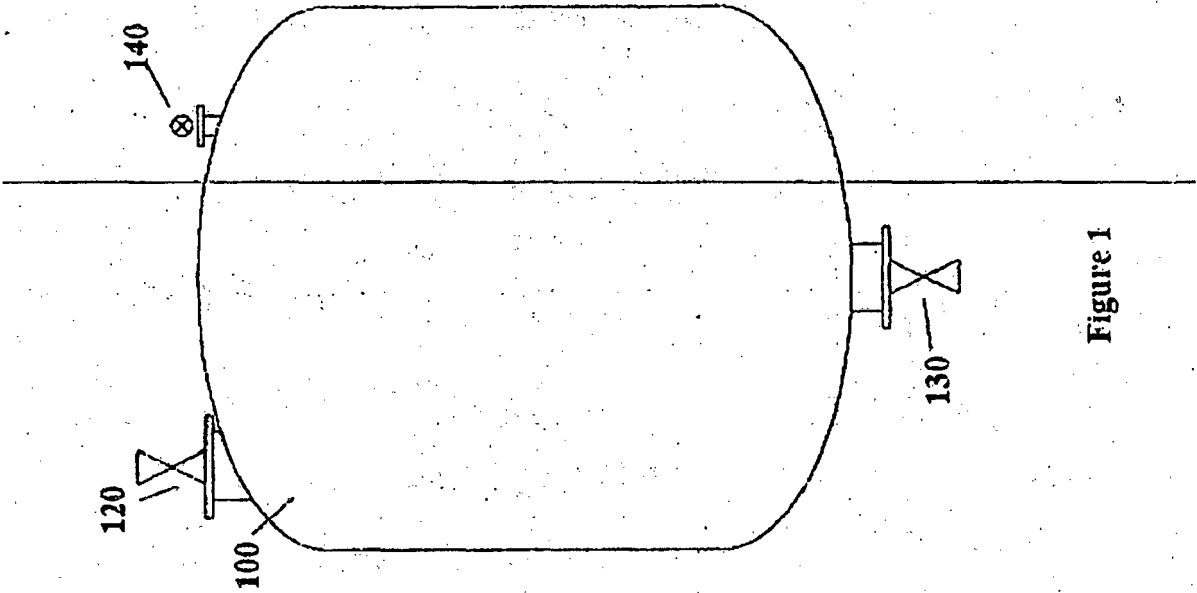


Figure 1

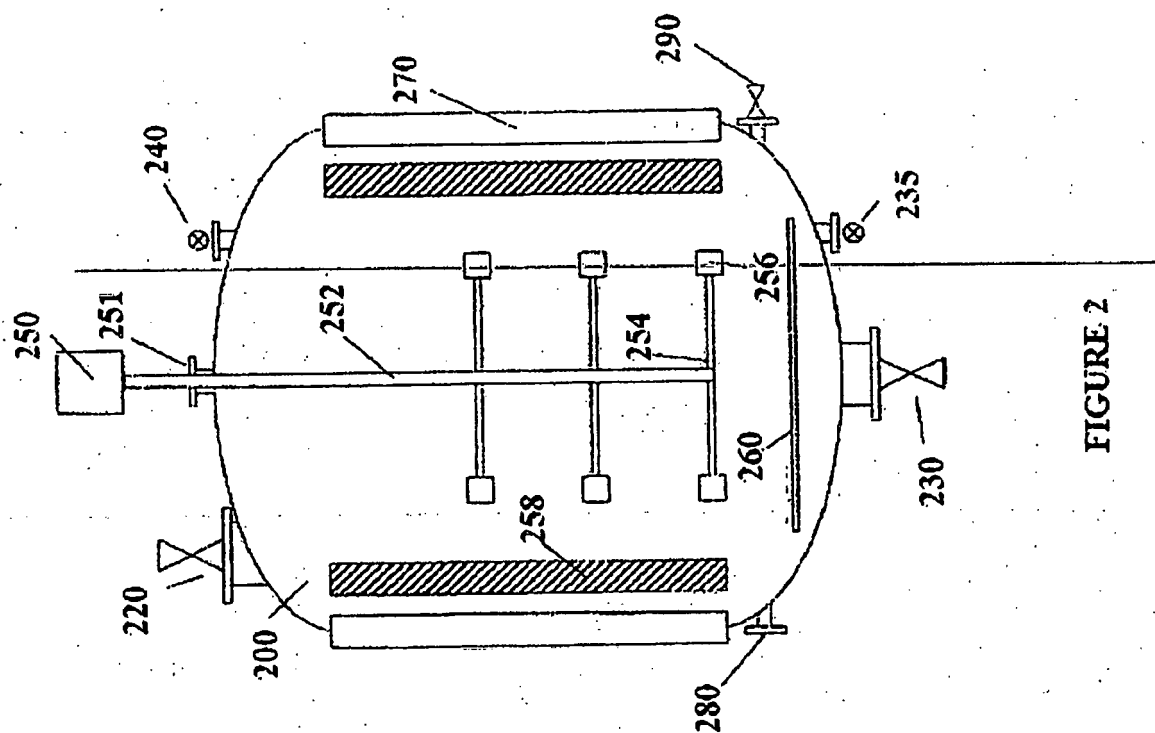


FIGURE 2

Figure 3
Mevalonate pathway

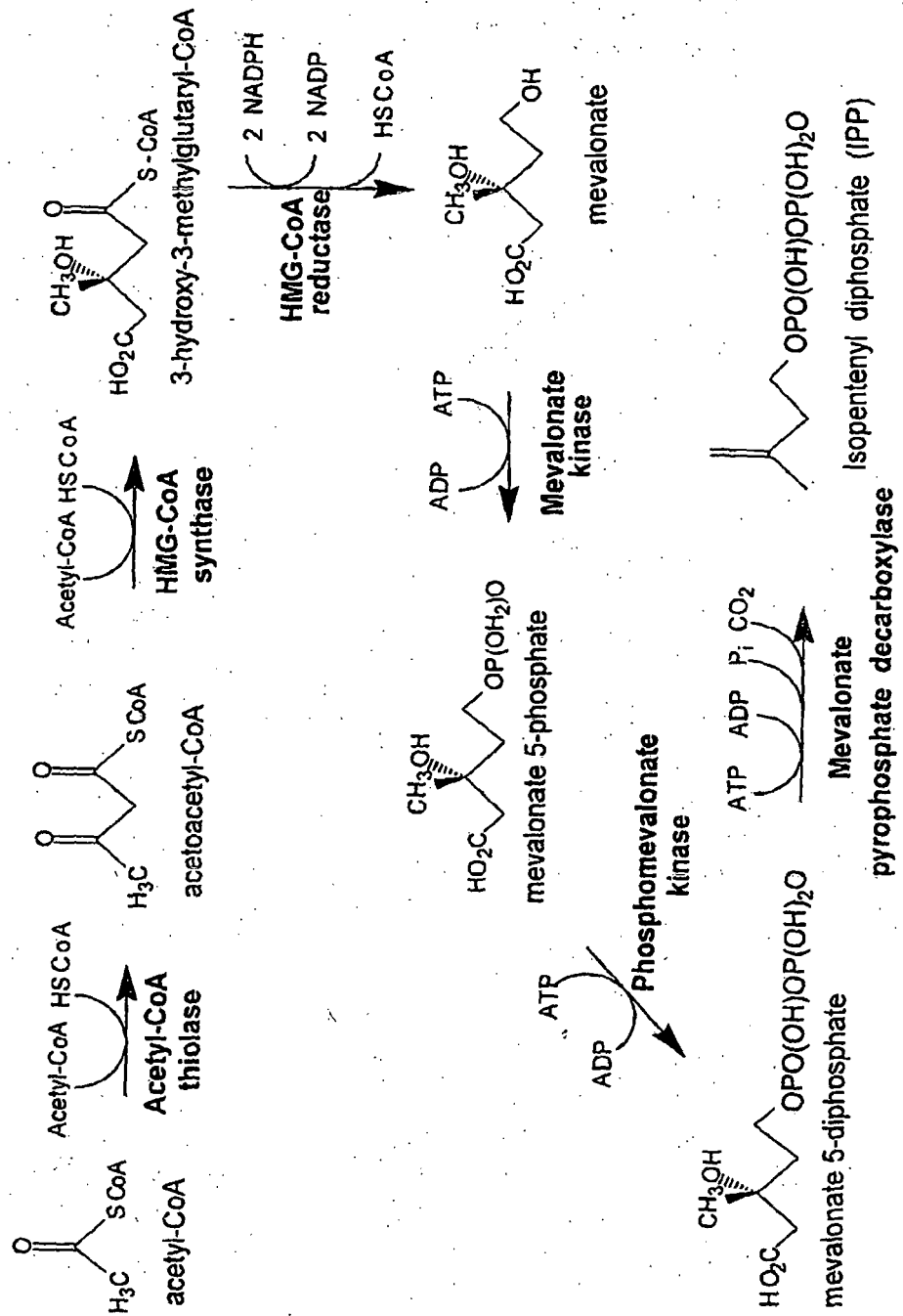


Figure 4
DXP pathway

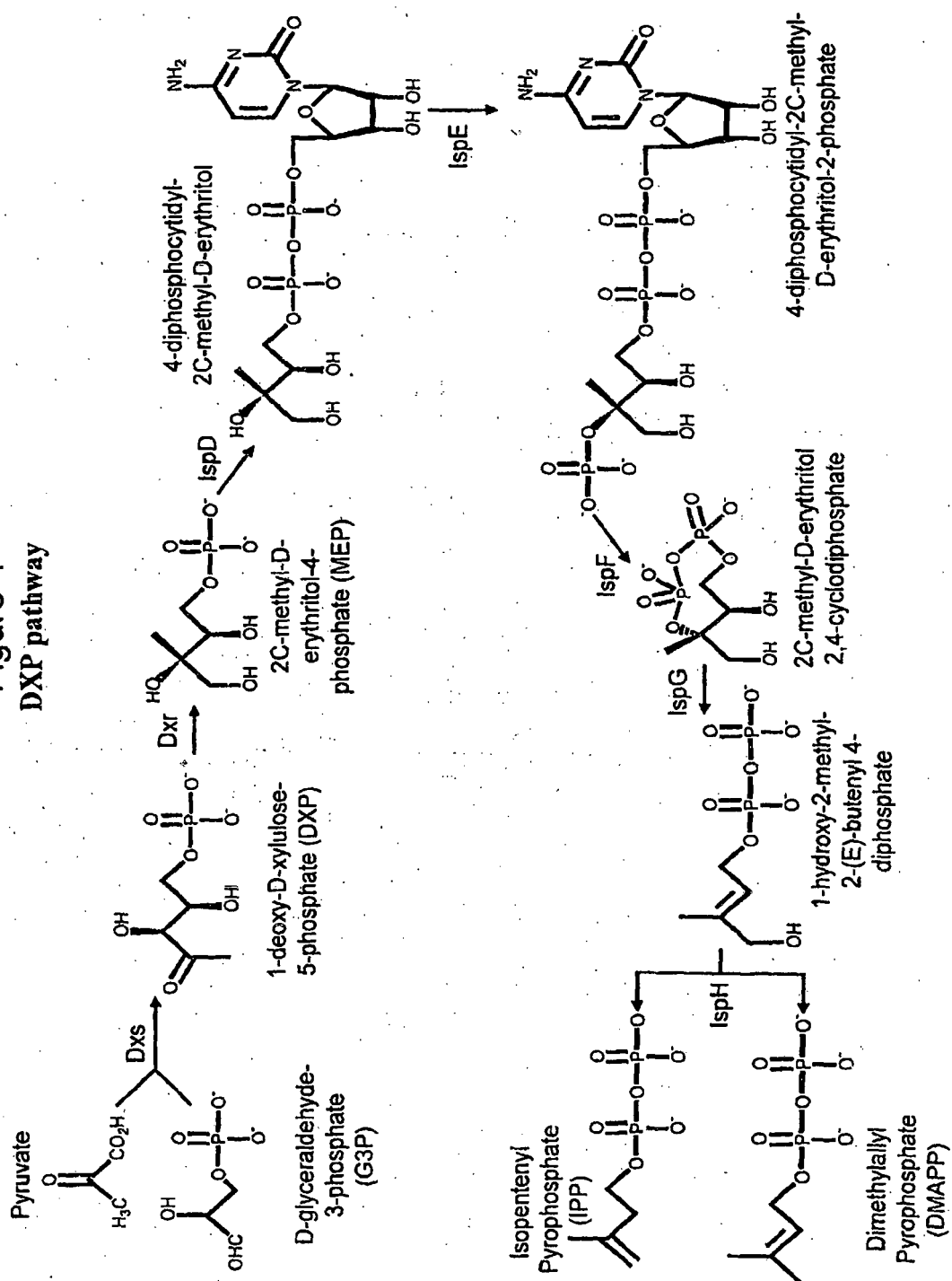


Figure 5

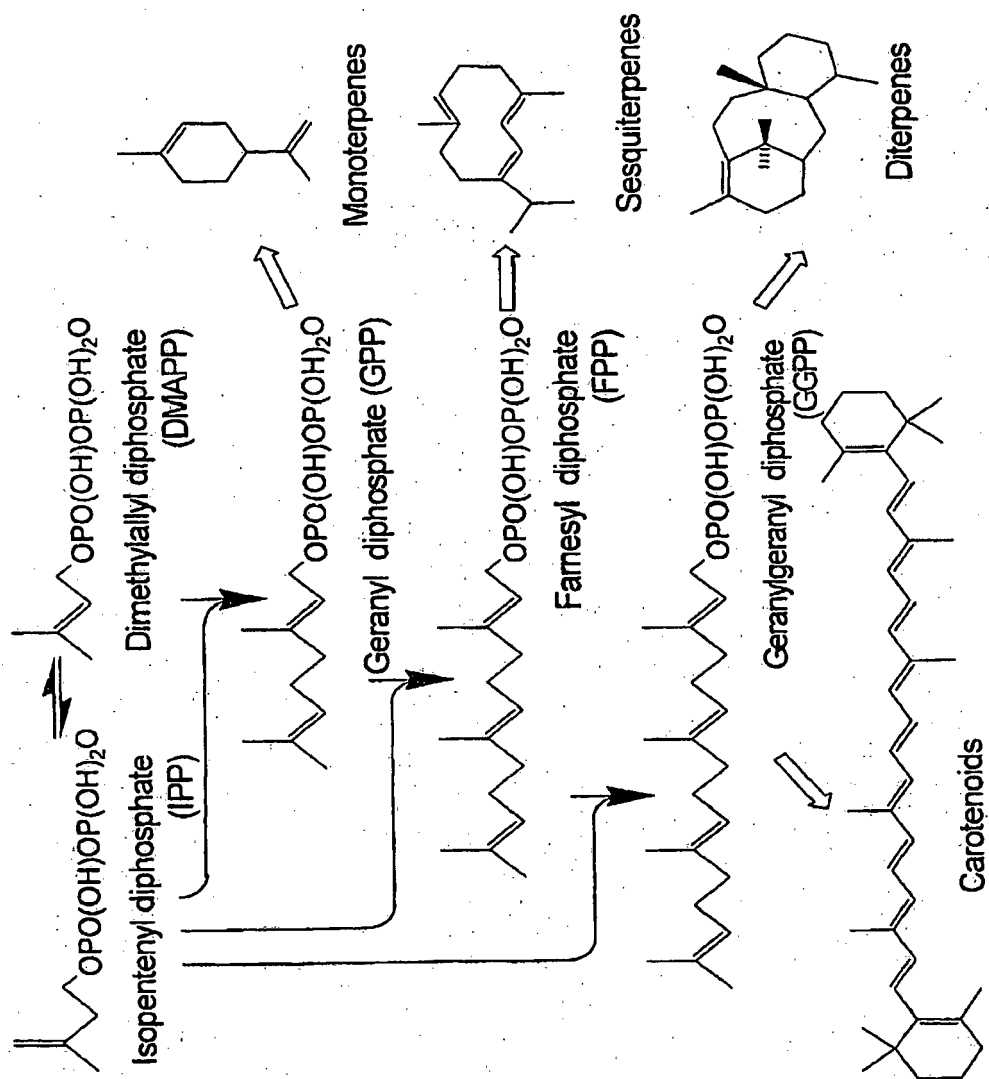


Figure 6

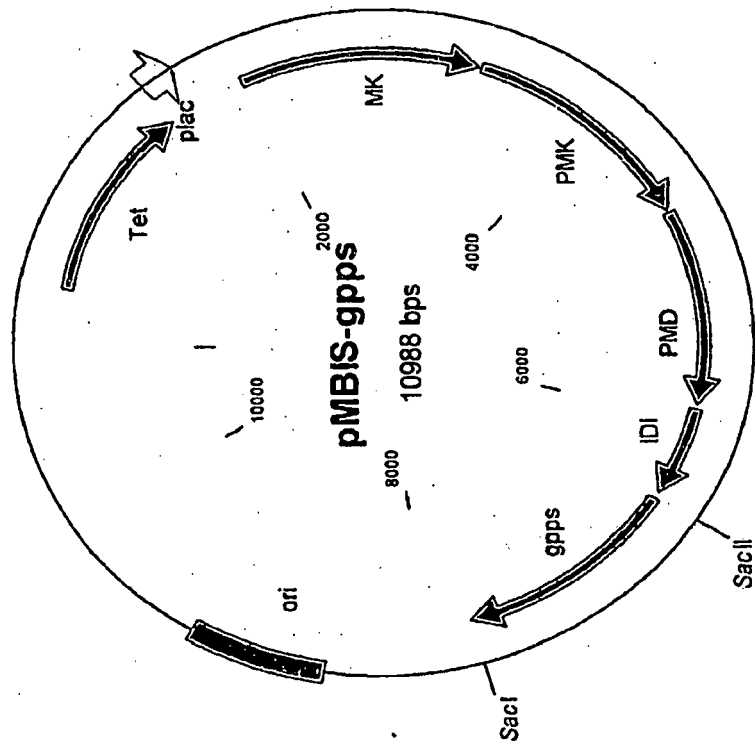


Figure 7

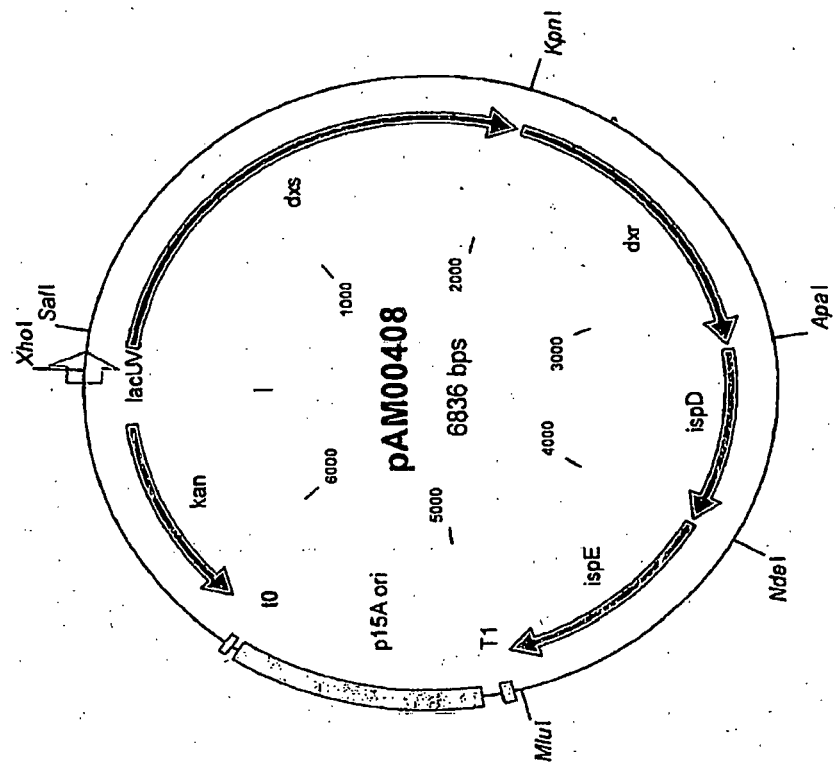


Figure 8

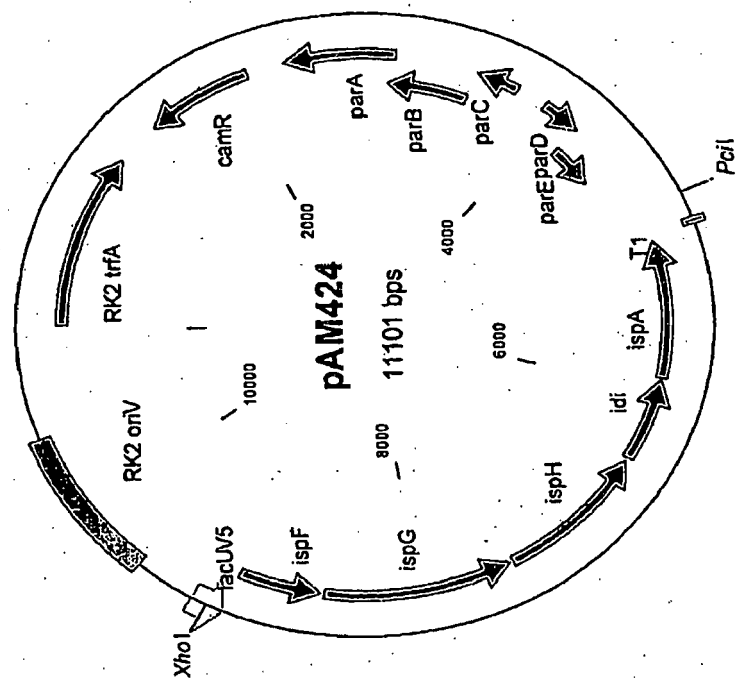
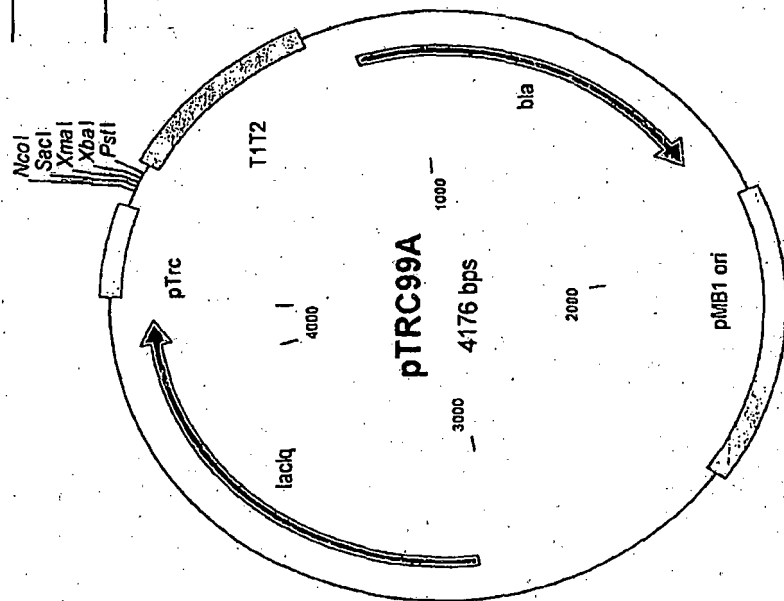
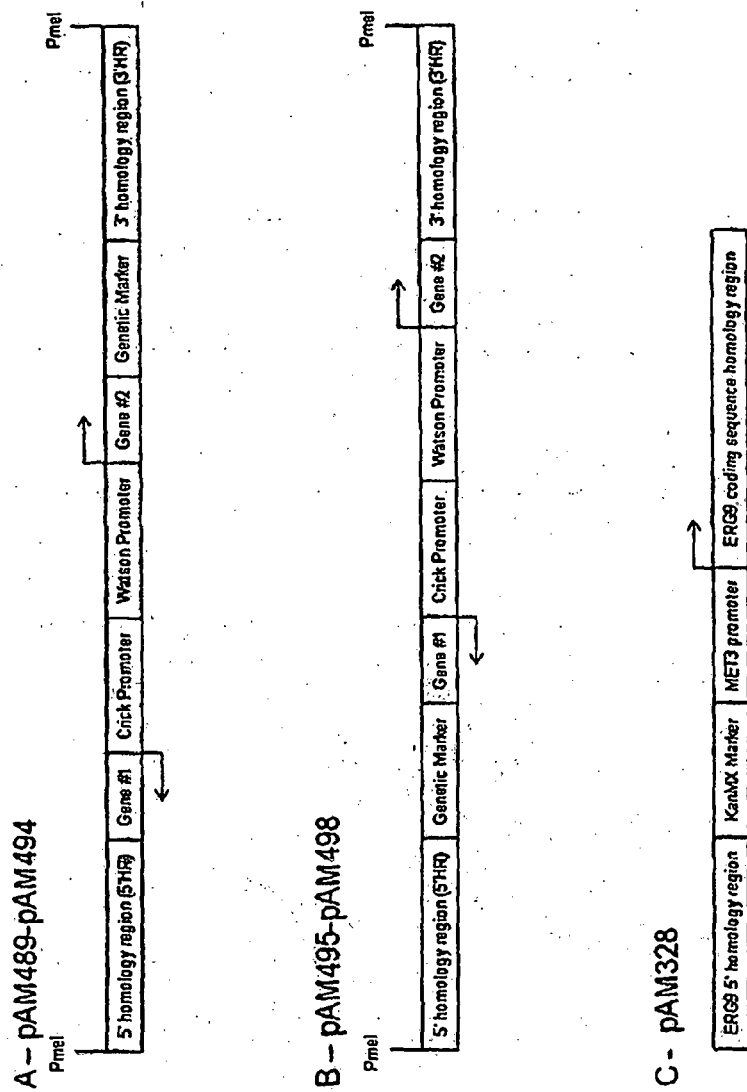


Figure 9



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 10



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 20040005678 A [0123]
- US 4500707 A [0144]
- US 5436327 A [0144]
- US 5700637 A [0144]
- US 20050084972 A [0147]
- US 20040131637 A [0150]
- WO 9617951 A [0150]
- US 7192751 B [0170] [0174] [0175] [0196]
- WO 2007005604 PCT [0252] [0276]

Non-patent literature cited in the description

- **NEWMAN et al.** *Biotechnology and Bioengineering*, 28 July 2006, vol. 95, 684-691 [0001]
- **PECHOUUS et al.** *Planta*, 2004, vol. 219 (1), 84-94 [0125]
- **PICAUD et al.** *Phytochemistry*, 2005, vol. 66 (9), 961-967 [0126]
- **SONG, L.** *Applied Biochemistry and Biotechnology*, 2006, vol. 128, 149-158 [0127]
- **BENNETZEN ; HALL.** *J. Biol. Chem.*, 1982, vol. 2757 (6), 3026-3031 [0142]
- **GOUY ; GAUTIER.** *Nucleic Acids Res.*, 1982, vol. 10 (22), 7055-7074 [0142]
- **EYRE-WALKER.** *Mol. Biol. Evol.*, 1996, vol. 13 (6), 864-872 [0142]
- **NAKAMURA et al.** *Nucleic Acids Res.*, 2000, vol. 28 (1), 292 [0142]
- **MATTEUCI et al.** *Tet. Lett.*, 1980, vol. 521, 719 [0144]
- **SIKORSKI ; HIETER.** *Genetics*, 1989, vol. 122, 19-27 [0147]
- **CHRISTAINSON et al.** *Gene*, 1992, vol. 110, 119-122 [0147]
- **BRUSCHI ; LUDWIG.** *Curr. Genet.*, 1988, vol. 15, 83-90 [0147]
- **ERHANRT et al.** *J. Bacteriol.*, 1983, vol. 156 (2), 625-635 [0147]
- **OKKELS.** *Annals of the New York Academy of Sciences*, 1996, vol. 782 (1), 202-207 [0147]
- **PULKKINEN ; MILLER.** *J. Bacteriol.*, 1991, vol. 173 (1), 86-93 [0150]
- **ALPUCHE-ARANDA et al.** *Proc. Natl. Acad. Sci. USA.*, 1992, vol. 89 (21), 10079-83 [0150]
- **HARBOME et al.** *Mol. Micro.*, 1992, vol. 6, 2805-2813 [0150]
- **DUNSTAN et al.** *Infect. Immun.*, 1999, vol. 67, 5133-5141 [0150]
- **MCKELVIE et al.** *Vaccine*, 2004, vol. 22, 3243-3255 [0150]
- **CHATFIELD et al.** *Biotechnol.*, 1992, vol. 10, 888-892 [0150]
- **SHETRON-RAMA et al.** *Infect. Immun.*, 2002, vol. 70, 1087-1096 [0150]
- **VALDIVIA ; FALKOW.** *Mol. Microbiol.*, 1996, vol. 22, 367-378 [0150]
- Topics in Molecular and Structural Biology. **HILLEN et al.** *Protein-Nucleic Acid Interaction*. 1989, vol. 10, 143-162 [0150]
- **MELTON et al.** *Nucl. Acids Res.*, 1984, vol. 12, 7035-7056 [0150]
- **GUZMAN et al.** *J. Bacteriol.*, 1995, vol. 177, 4121-4130 [0153]
- **KIM et al.** *Gene*, 1996, vol. 181, 71-76 [0153]
- **HOFFMANN et al.** *FEMS Microbiol Lett.*, 1999, vol. 177 (2), 327-34 [0153]
- **WEST.** *Mol. Cell. Biol.*, 1984, vol. 4 (11), 2467-2478 [0154]
- **DEBOER et al.** *Proc. Natl. Acad. Sci. U.S.A.*, 1983, vol. 80, 21-25 [0155]
- **SHINE et al.** *Nature*, 1975, vol. 254, 34 [0156]
- **STEITZ.** *Biological Regulation and Development: Gene Expression*. Plenum Publishing, 1979, vol. 1, 349 [0156]
- **LOOMAN et al.** *Nuc. Ac. Res.*, 1993, vol. 21, 4268-4271 [0156]
- **YUN.** *Mol. Microbiol.*, 1996, vol. 19, 1225-1239 [0156]
- **BITTER et al.** *Methods in Enzymology*, 1987, vol. 153, 516-544 [0157]
- **SIA.** *J. Bacteriol.*, 1995, vol. 177, 2789-97 [0160]
- **PANSEGRAU et al.** *J. Mol. Biol.*, 1994, vol. 239, 623-63 [0160]
- **GUZMAN et al.** *J. Bacteriol.*, 1995, vol. 177 (14), 4121-4130 [0170]
- **LUTZ ; BUJARD.** *Nucl Acids Res.*, 1997, vol. 25, 1203-1210 [0172] [0188] [0193] [0194] [0197]
- **KOVACH et al.** *Gene*, 1995, vol. 166 (1), 175-175 [0173]
- **KOVACH et al.** *Gene*, 1995, vol. 166 (1), 175-176 [0174]

- **GEISER et al.** *Bio Techniques*, 2001, vol. 31, 88-92 [0180]
- **ROBERTS et al.** *J Bacteriol.*, 1990, vol. 172 (11), 6204-6216 [0184]
- **MEYER et al.** *Science*, 1975, vol. 190, 1226-1228 [0194]
- **ROBERTS et al.** *J Bacteriol.*, 1990, vol. 172, 6204-6216 [0194]
- **MERKE et al.** *Ach. Biochem. Biophys.*, 2000, vol. 381, 173-180 [0196]
- **AMMAN et al.** *Gene*, 1985, vol. 40, 183-190 [0196] [0200]
- **AMANN et al.** *Gene*, 1988, vol. 69, 301-3115 [0198]
- **MUMBERG.** *Nucl. Acids. Res.*, 1994, vol. 22 (25), 5767-5768 [0201]
- **BRADFORD.** *Anal Biochem.*, 1976, vol. 72, 248-254 [0213]
- **MARTIN et al.** *Biotechnol. Bioeng.*, 2001, vol. 75, 497-503 [0219]
- **RO et al.** *Nature*, 2006, vol. 440, 40-943 [0252]
- **Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual** [0252]
- **ERHART ; HOLLENBERG.** *J. Bacteriol.*, 1983, vol. 156, 625-635 [0252]
- **J. P. VAN DIJKEN et al.** *Enzyme Microb Technol*, 01 June 2000, vol. 26, 706 [0258]
- **D. ROSE ; F. WINSTON ; P. HEITER.** *Methods in yeast genetics: a laboratory course manual.* Cold Spring Harbor Laboratory Press, 1990 [0258]
- **R. H. SCHIESTL ; R. D. GIETZ.** *Curr Genet*, December 1989, vol. 16, 339 [0259]
- **M. S. LONGTINE et al.** *Yeast*, July 1998, vol. 14, 953 [0259]
- **SCHIESTL ; GIETZ.** *Curr. Genet.*, 1989, vol. 16, 339 [0260]
- **RO et al.** *Nature*, 2006, vol. 440, 940-943 [0276]