METHODS OF DEVELOPING TERPENE SYNTHASE VARIANTS

1. FIELD OF THE INVENTION

The present disclosure relates to methods of developing terpene synthase variants through engineered host cells. Particularly, the disclosure provides methods of developing terpene synthase variants with improved *in vivo* performance that are useful in the commercial production of terpene products. Further encompassed in the present disclosure are superior terpene synthase variants, and host cells comprising such terpene synthase variants.

2. BACKGROUND

Terpenes are a large class of hydrocarbons that are produced in many organisms. They are derived by linking units of isoprene (C_5H_8), and are classified by the number of isoprene units present. Hemiterpenes consist of a single isoprene unit. Isoprene itself is considered the only hemiterpene. Monoterpenes are made of two isoprene units, and have the molecular formula $C_{10}H_{16}$. Examples of monoterpenes are geraniol, limonene, and terpineol. Sesquiterpenes are composed of three isoprene units, and have the molecular formula $C_{15}H_{24}$. Examples of sesquiterpenes are farnesenes, farnesol and patchoulol. Diterpenes are made of four isoprene units, and have the molecular formula $C_{20}H_{32}$. Examples of diterpenes are cafestol, kahweol, cembrene, and taxadiene. Sesterterpenes are made of five isoprene units, and have the molecular formula $C_{25}H_{40}$. An example of a sesterterpenes is geranylfarnesol. Triterpenes consist of six isoprene units, and have the molecular formula $C_{30}H_{48}$. Tetraterpenes contain eight isoprene units, and have the molecular formula $C_{40}H_{64}$. Biologically important tetraterpenes include the acyclic lycopene, the monocyclic gamma-carotene, and the bicyclic alpha- and beta-carotenes. Polyterpenes consist of long chains of many isoprene units. Natural rubber consists of polyisoprene in which the double bonds are cis.

When terpenes are chemically modified *(e.g.,* via oxidation or rearrangement of the carbon skeleton) the resulting compounds are generally referred to as terpenoids, which are also known as isoprenoids. Isoprenoids play many important biological roles, for example, as quinones in electron transport chains, as components of membranes, in subcellular targeting and regulation via protein prenylation, as photosynthetic pigments including carotenoids, chlorophyll, as hormones and cofactors, and as plant defense

compounds with various monoterpenes, sesquiterpenes, and diterpenes. They are industrially useful as antibiotics, hormones, anticancer drugs, insecticides, and chemicals.

Terpenes are biosynthesized through condensations of isopentenyl pyrophosphate (isopentenyl diphosphate or IPP) and its isomer dimethylallyl pyrophosphate (dimethylallyl diphosphate or DMAPP). Two pathways are known to generate IPP and DMAPP, namely the mevalonate-dependent (MEV) pathway of eukaryotes, and the mevalonate-independent or deoxyxylulose-5-phosphate (DXP) pathway of prokaryotes. Plants use both the MEV pathway and the DXP pathway. IPP and DMAPP in turn are condensed to polyprenyl diphosphates (*e.g.*, geranyl disphosphate or GPP, farnesyl diphosphate or FPP, and geranylgeranyl diphosphate or GGPP) through the action of prenyl disphosphate synthases (*e.g.*, GPP synthase, FPP synthase, and GGPP synthase, respectively).

The polyprenyl diphosphate intermediates are converted to more complex isoprenoid structures by terpene synthases. Terpene synthases are organized into large gene families that form multiple products. Examples of terpene synthases include sesquiterpene synthases, which convert FPP into sesquiterpenes. An example of a sesquiterpene synthase is farnesene synthase, which converts FPP to farnesene. The reaction mechanism of terpene synthases has been extensively investigated and is well understood. Overall, three steps are required to convert a diphosphate substrate such as FPP to its isoprenoid product: a) formation of enzyme-substrate complex (ES), b) formation of an enzyme-bound reactive carbocation intermediate, subsequent rearrangements, and the formation of product (EP), and c) release of product from the enzyme-product complex. *In vitro* kinetic and pre-steady state kinetic studies on terpene synthase catalyzed reactions have shown that the overall rate-limiting step for the reactions is the release of product (Cane *et al.* (1997) *Biochemistry,* 36(27):8332-9, and Mathis *et al.* (1997) *Biochemistry* 36(27):8340-8). The turnover rates of terpene synthases are low, generally measured at less than 0.5 per second (Cane, D. C. (1990) *Chem. Rev.* 90:1089-1103).

Terpene synthases are important in the regulation of pathway flux to an isoprenoid because they operate at metabolic branch points and often compete with other metabolic enzymes for a prenyl diphosphate pool. For example, FPP is the precursor to many cellular molecules including squalene, dolichols, and the cofactor heme. In engineered microbes where the production of sesquiterpenes such as farnesene is desired, the terpene synthases hold the key to high yield production of such terpenes. However, because they are slow

enzymes, terpene synthases are often the bottlenecks in the metabolic pathways. In addition, they can suffer from other shortcomings such as substrate inhibition that limit the kinetic capacity required for efficient production of terpenes in engineered microbial hosts (Crock *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:12833-12838).

Hence, there are potentially enormous benefits to improving the catalytic efficiency of terpene synthases so that these enzymes would no longer limit the overall metabolic flux to an isoprenoid. Attempts to engineer terpene synthases for altered product specificity as well as the use of rational approaches such as those based on structural guidance or adaptive evolution have been described previously (Greenhagen *et al.* (2006) *Proc. Natl. Acad. Sci. USA* 103:9826-9831; O'Maille *et al.* (2008) *Nat. Chem. Biol.* 4:617-623; Yoshikuni *et al.* (2006) *Nature* 440:1078-1082; Yoshikuni *et al.* (2008) *Chem. Biol.* 15:607-618). However, these studies have fallen short of improving the kinetic capacity of terpene synthases while also maintaining their product specificity. In addition, the application of conventional protein engineering strategies, such as directed evolution, has been devoid for terpene synthases primarily because of the lack of available and effective high throughput screening methods (Yoshikuni *et al.* (2008) (*supra*)). There thus remains a need for reliable and high throughput methods for improving the catalytic efficiency of terpene synthases, and for terpene synthase variants that have such improved catalytic efficiency.

3. SUMMARY OF THE INVENTION

The present disclosure relates to methods of developing terpene synthase variants through engineered host cells. Particularly, the disclosure provides methods of developing terpene synthase variants with improved *in vivo* performance. The methods also allow for the continued improvement of the *in vivo* performance of these enzymes.

In one aspect, the present invention provides a screening method for a sesquiterpene synthase variant with improved *in vivo* performance, comprising the steps of:

a) engineering a host cell expressing a control sesquiterpene synthase to comprise an elevated level of FPP, wherein the elevated level of FPP reduces the viability of the host cell compared to a parent cell not comprising the elevated level of FPP;

b) expressing in the host cell a test sesquiterpene synthase instead of the control sesquiterpene synthase, wherein the test sesquiterpene synthase is a variant of the control sesquiterpene synthase; and

c) identifying the test sesquiterpene synthase as having improved *in vivo* performance compared to the control sesquiterpene synthase by an increase in viability of the host cell expressing the test sesquiterpene synthase compared to the host cell expressing the control terpene synthase.

In some embodiments, the host cell is plated on an agar plate, and a host cell comprising a test terpene synthase variant with improved *in vivo* performance is identified by colony growth. In some embodiments, the method further comprises selecting and/or isolating the test sesquiterpene synthase having improved *in vivo* performance.

In some embodiments, a collection of sesquiterpene synthase variants is expressed in a collection of host cells. In some embodiments, the collection of sesquiterpene synthase variants comprises from 2 to 5, from 5 to 10, from 10 to 50, from 50 to 100, from 100 to 500, from 500 to 1,000, from 1,000 to 10,000, from 10,000 to 100,000, from 100,000 to 1,000,000, and more, sesquiterpene synthase variants.

In some embodiments, the screening method is used in an iterative fashion, wherein the test sesquiterpene synthase identified in an iteration is used as the control sesquiterpene synthase of the next iteration, and wherein the host cell in an iteration comprises such elevated level of FPP that it has reduced viability in the presence of the test sesquiterpene synthase identified in the previous iteration compared to a parent cell not comprising the elevated level of FPP.

In another aspect, provided herein is a composition comprising two cell subpopulations derived from a common population of host cells comprising an elevated level of FPP, wherein:

a) the first subpopulation comprises a control sesquiterpene synthase, wherein the elevated level of FPP reduces the viability of cells of the first subpopulation compared to the viability of a parent cell not comprising the elevated level of FPP; and

b) the second subpopulation comprises a test sesquiterpene synthase, wherein the test sesquiterpene synthase is a variant of the control sesquiterpene synthase.

In some embodiments, the viability of the cells of the second subpopulation is greater than the viability of the cells of the first subpopulation.

In another aspect, the present invention provides a second screening method for identifying terpene synthase variants with improved *in vivo* performance, comprising the steps of:

a) providing a host cell expressing a control terpene synthase and having a growth rate;

b) expressing in the host cell a test terpene synthase instead of the control terpene synthase, wherein the test terpene synthase is a variant of the control terpene synthase; and

d) identifying the test terpene synthase as having improved *in vivo* performance compared to the control terpene synthase by a decreased growth rate of the host cell expressing the test terpene synthase compared to the growth rate of the host cell expressing the control terpene synthase.

In yet another aspect, the present invention provides a competition method for identifying and/or ranking the *in vivo* performance of terpene synthase variants, comprising the steps of:

a) dividing a population of host cells into a control population and a test population;

b) expressing in the control population a control terpene synthase and a comparison terpene synthase, wherein the control terpene synthase can convert a polyprenyl diphosphate to a first terpene, and wherein the comparison terpene synthase can convert a polyprenyl diphosphate to a second terpene;

c) expressing in the test population the comparison terpene synthase and a test terpene synthase, wherein the test terpene synthase is a variant of the control terpene synthase, and wherein the comparison terpene synthase is expressed at similar levels in the test population and in the control population; and

d) measuring a ratio of the first terpene over the second terpene in the test population and in the control population.

In separate embodiments, the competition method is applied to identify and/or to rank terpene synthases selected from the group consisting of monoterpene synthases, diterpene synthases, sesquiterpene synthases, sesterterpene synthases, triterpene synthases, tetraterpene synthases, and polyterpene synthases.

In some embodiments, the competition method is used to screen a library of mutant terpene synthases on the basis that compared to the control terpene synthase, a terpene synthase variant with improved *in vivo* performance is capable of diverting more flux from a polyprenyl diphosphate substrate to its terpene product, thus, giving a higher ratio of terpene of interest/comparison terpene (*i.e.*, first terpene/second terpene). In such embodiments, it is important that the test terpene synthase is expressed at a similar level in the test population as the control terpene synthase is expressed in the control population.

In other embodiments, the competition method is used to identify a promoter of a desired strength. In such embodiments, the control terpene synthase and the test terpene

synthase are identical, and the control population and test population differ in the expression level of the control terpene synthase.

In another aspect, provided herein is a composition comprising two cell subpopulations derived from a common population of host cells, wherein:

a) the first subpopulation comprises a control terpene synthase and a comparison terpene synthase, wherein the control terpene synthase converts a polyprenyl diphosphate to a first terpene, and wherein the comparison sesquiterpene synthase converts the polyprenyl diphosphate to a second terpene; and

b) the second subpopulation comprises a test terpene synthase and the comparison terpene synthase, wherein the control terpene synthase converts the polyprenyl diphosphate to the first terpene, and wherein the test terpene synthase is a variant of the control terpene synthase.

In some embodiments, the ratio of the first terpene over the second terpene is greater in the second subpopulation compared to that in the first subpopulation.

In yet another aspect, provided herein are isolated β–farnesene synthase variants, and isolated nucleic acids comprising a nucleotide sequence encoding such β–farnesene synthase variants, having an amino acid sequence as given in SEQ ID NO: 111 but comprising one or more amino acid substitutions at positions selected from the group consisting of positions 2, 3, 4, 6, 9, 11, 18, 20, 24, 35, 38, 50, 61, 72, 80, 89, 105, 115, 144, 196, 211, 251, 280, 288, 319, 348, 357, 359, 369, 371, 385, 398, 423, 433, 434, 442, 444, 446, 460, 467, 488, 495, 505, 526, 531, 556, 572, and 575 of SEQ ID NO: 111.

In yet another aspect, the present invention provides a genetically modified host cell that comprises:

(a) a heterologous β–farnesene synthase, wherein the heterologous β–farnesene synthase is a variant of a β–farnesene synthase encoded by SEQ ID NO: 111; and

(b) a MEV pathway or DXP pathway enzyme;

wherein the host cell makes at least 15% more of a β–farnesene compared to a parent cell that comprises the MEV pathway or DXP pathway enzyme and the β–farnesene synthase encoded by SEQ ID NO: 111.

In yet another aspect, provided herein is a method of producing β–farnesene comprising the steps of:

(a) obtaining a plurality of genetically modified host cells comprising:

i) a first heterologous nucleotide sequence encoding a variant of a β–farnesene synthase encoded by SEQ ID NO: 111; and

ii) a second heterologous nucleotide sequence encoding a MEV pathway or DXP pathway enzyme;

(b) culturing said genetically modified host cells in a medium comprising a carbon source under conditions suitable for making the β–farnesene; and

(c) recovering the β–farnesene from the medium.

4. BRIEF DESCRIPTION OF THE FIGURES

The present disclosure is best understood when read in conjunction with the accompanying figures, which serve to illustrate the preferred embodiments. It is understood, however, that the disclosure is not limited to the specific embodiments disclosed in the figures.

Figure 1A-Z provides maps of several chromosomal integration constructs used in the generation of host cells of the invention.

Figure 2 provides an image of several agar plates on which were plated *Escherichia coli* host cells comprising active and inactive sesquiterpene synthases for FPP starvation-based selection.

Figure 3 provides an image of two agar plates on which were plated *Escherichia coli* host cells comprising active and inactive sesquiterpene synthases for FPP toxicity-based growth selection.

Figure 4 provides farnesene titers obtained by GC analysis of *Escherichia coli* host cells comprising various farnesene synthase coding sequences.

Figure 5 provides an image of agar plates on which were plated *Saccharomyces cerevisiae* host cells comprising active and inactive sesquiterpene synthases for FPP toxicity-based growth selection.

Figure 6 provides farnesene titers obtained by Nile Red fluorescence analysis of *Saccharomyces cerevisiae* host cells comprising either chromosomally intergrated or extrachromosomally maintained farnesene synthase coding sequences.

Figure 7 provides farnesene titers obtained by GC analysis of *Saccharomyces cerevisiae* host cells ranked by sesquiterpene synthase competition.

Figure 8 provides farnesene/trichodiene titer ratios obtained by GC analysis of *Saccharomyces cerevisiae* host cells comprising increasing copy numbers of farnesene synthase coding sequences.

Figure 9 provides a comparison of farnesene titers obtained by GC analysis versus Nile Red fluorescence analysis of *Escherichia coli* host cells of a sesquiterpene synthase library.

Figure 10 provides farnesene titers obtained by GC analysis of *Escherichia coli* host strains identified from a library of FS variants screened by Nile Red fluorescence.

Figures 11 provides farnesene titers obtained by Nile Red fluoresence (A) and GC analysis (B) of *Saccharomyces cerevisiae* host strains identified from a library of FS variants screened by FPP toxicity based growth selection.

Figure 12 provides farnesene titers obtained by Nile Red fluourescence analysis of *Saccharomyces cerevisiae* host strains identified from a library of FS variants by FPP toxicity-based growth selection.

Figure 13 provides maps of various expression plasmids used in the generation of host cells of the invention.

Figure 14 provides farnesene titers obtained by GC analysis of *Saccharomyces cerevisiae* host strains comprising single chromosomally integrated copies of FS variant coding sequences.

Figure 15 provides a schematic representation of the MEV pathway for the production of IPP and DMAPP.

Figure 16 provides a schematic representation of the DXP pathway for the production of IPP and DMAPP.

Figure 17 provides amorphadiene/trichodiene titer ratios obtained by GC analysis of *Saccharomyces cerevisiae* host cells comprising coding sequences for amorphadiene synthase variants.

Figure 18 provides limonene/myrcene titer ratios obtained by GC analysis of *Saccharomyces cerevisiae* host cells comprising coding sequences for limonene synthase variants.

5. DETAILED DESCRIPTION OF THE EMBODIMENTS

5.1. Definitions

The following terms used herein shall have the meanings as indicated below.

As used herein, the term "terpene synthase variant" refers to a terpene synthase that compared to a selected terpene synthase has a different nucleotide or amino acid sequence. For example, compared to the wild-type sequence of the selected terpene synthase, the terpene synthase variant may comprise nucleotide additions, deletions, and/or substitutions that may or may not result in changes to the corresponding amino acid

sequence. In some embodiments where nucleotide changes do not result in changes to the amino acid sequence, the changes may nonetheless effect improved activity of the synthase, for example, through codon optimization. In other embodiments, the terpene synthase variant comprises amino acid additions, deletions, and/or substitutions. Accordingly, as used herein, the term "sesquiterpene synthase variant" refers to a sesquiterpene synthase that compared to a selected sesquiterpene synthase has a different nucleotide or amino acid sequence. For example, compared to the selected sesquiterpene synthase, the sesquiterpene synthase variant may comprise nucleotide additions, deletions, and/or substitutions that may or may not result in changes to the corresponding amino acid sequence. In other embodiments, the terpene synthase variant comprises amino acid additions, deletions and/or substitutions.

As used herein, the term "engineered host cell" refers to a host cell that is generated by genetically modifying a parent cell using genetic engineering techniques (*i.e.*, recombinant technology). The engineered host cell may comprise additions, deletions, and/or modifications of nucleotide sequences to the genome of the parent cell.

As used herein, the term "heterologous" refers to what is not normally found in nature. The term "heterologous nucleotide sequence" refers to a nucleotide sequence not normally found in a given cell in nature. As such, a heterologous nucleotide sequence may be: (a) foreign to its host cell (*i.e.,* is "exogenous" to the cell); (b) naturally found in the host cell (*i.e.,* "endogenous") but present at an unnatural quantity in the cell (*i.e.,* greater or lesser quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus.

As used herein, the term "naturally occurring" refers to what is found in nature. For example, a terpene synthase 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 a naturally occurring terpene synthase. Conversely, as used herein, the term "naturally not occurring" refers to what is not found in nature but is created by human intervention.

As used herein, the term "biosynthetic enzyme" refers to an enzyme that functions in a biosynthetic pathway leading to the production of a naturally occurring molecule.

As used herein, the term "*in vivo* performance" refers to the ability of a terpene synthase to convert a polyprenyl diphosphate substrate to a terpene when expressed in a host cell. Accordingly, the term "improved *in vivo* performance" refers to an increased ability of a terpene synthase to convert a polyprenyl diphosphate substrate to a terpene when expressed in a host cell.

As used herein, the term "parent cell" refers to a cell that has an identical genetic background as a host cell disclosed herein except that it does not comprise the elevated intracellular level of FPP or does not comprise a particular heterologous nucleotide sequence, and that serves as the starting point for introducing said elevated intracellular level of FPP or said heterologous nucleotide sequence leading to the generation of a host cell disclosed herein.

5.2. General Overview

The present disclosure relates to methods of developing terpene synthase variants through engineered host cells. Particularly, the disclosure provides methods of developing terpene synthase variants with improved *in vivo* performance. The methods also allow for the continued improvement of the *in vivo* performance of these enzymes.

In one aspect, the present invention provides a screening method for terpene synthase variants with improved *in vivo* performance. In some embodiments, terpene synthase variants with improved *in vivo* performance are identified by their ability to rescue engineered host cells from cell death. The engineered host cells comprise genetic modifications that cause elevated intracellular levels of FPP. Because FPP is highly toxic to cells and thus reduces cell viability (Withers *et al.* (2007) *Appl. Environ. Microbiol.* 73:6277-6283), to achieve a viability that is comparable to that of a parent cell that does not comprise the elevated level of intracellular FPP, the engineered host cells require a sufficiently active sesquiterpene synthase to reduce the intracellular levels of FPP.

The presently provided screening method thus comprises the following steps:

a) engineering a host cell expressing a control sesquiterpene synthase to comprise an elevated level of FPP, wherein the elevated level of FPP reduces the viability of the host cell compared to a parent cell not comprising the elevated level of FPP;

b) expressing in the host cell a test sesquiterpene synthase instead of the control sesquiterpene synthase, wherein the test sesquiterpene synthase is a variant of the control sesquiterpene synthase; and

c) identifying the test sesquiterpene synthase as having improved *in vivo* performance compared to the control sesquiterpene synthase by an increase in viability of the host cell expressing the test sesquiterpene synthase compared to the host cell expressing the control sesquiterpene synthase.

In some embodiments, the method further comprises selecting and/or isolating the test sesquiterpene synthase having improved *in vivo* performance.

It is most convenient if the elevated level of FPP in the host cell is inducible. Induction may occur in response to an inducing agent or specific growth conditions such as, for example, temperature. The elevated level of FPP in the host cell may range from about 10% to at least about 1,000-fold, or more, higher than the level of FPP of the parent cell.

The reduced viability of the host cell expressing the control sesquiterpene synthase compared to the parent cell may range from decreased cell growth to lethality. Thus, in some embodiments, the host cell expressing the control sesquiterpene synthase produces a reduced number of progeny cells in a liquid culture or on an agar plate compared to the parent cell. In other embodiments, the host cell expressing the control sesquiterpene synthase produces no progeny cells in a liquid culture or on an agar plate compared to the parent cell. Accordingly, the increase in viability of the host cell expressing the test sesquiterpene synthase instead of the control sesquiterpene synthase may be apparent in liquid culture by a higher number of progeny cells, or on an agar plate by a larger colony size, compared to the number of progeny cells or colony size produced by the host cell expressing the control sesquiterpene synthase.

The elevated level of FPP in the host cell may be effected by modifying the expression and/or activity of an enzyme involved in the production of FPP or its precursors in the host cell. In some such embodiments, the expression and/or activity of an enzyme of the MEV or DXP pathway is modified. In some such embodiments, the expression and/or activity of a HMG-CoA reductase and/or a mevalonate kinase is modified. Alternatively, the elevated level of FPP in the host cell may be effected by modifying the expression and/or activity of an enzyme involved in the utilization of FPP or its precursors in the host cell. In some such embodiments, the expression and/or activity of a squalene synthase is modified.

The control sesquiterpene synthase may be a naturally occurring sesquiterpene synthase or a naturally not occurring sesquiterpene synthase. The test sequiterpene synthase may differ from the control sesquiterpene synthase by comprising one or more amino acid substitutions, deletions, and/or additions. In addition or alternatively, the test sequiterpene synthase may comprise identical amino acids as the control sesquiterpene synthase but the codons encoding these amino acids may differ between the test sesquiterpene synthase and the control sesquiterpene synthase. In some such embodiments, the codons are optimized for usage in the host cell.

In some embodiments, the control sesquiterpene synthase is selected from the group consisiting of a β–farnesene synthase, an α–farnesene synthase, a trichodiene synthase, a patchoulol synthase, an amorphadiene synthase, a valencene synthase, a farnesol synthase, a nerolidol synthase, and a nootkatone synthase. In some such embodiments, the control sesquiterpene synthase is a β–farnesene synthase of *Artemisia annua*. In some such embodiments, the control sesquiterpene synthase has an amino acid sequence as given in SEQ ID NO: 111.

To be able to compare the viability of the host cell in the presence of the test sesquiterpene synthase to that of the host cell in the presence of the control sesquiterpene synthase, it is necessary to ensure similar expression levels of the control sesquiterpene synthase and the test sesquiterpene synthase in the host cell. This can be accomplished by placing the nucleotide sequences encoding the sesquiterpene synthases in the two host cells under the control of the same regulatory elements.

To prevent a competitive growth situation in which fast growing false positive host cells comprising a growth promoting mutation rather than an improved sesquiterpene synthase variant take over a host cell culture, one embodiment of the screening method involves an agar-plate based selection system. In this embodiment, the host cell is plated on an agar plate, and a host cell comprising a test sesquiterpene synthase variant with improved *in vivo* performance is identified by colony growth.

One major advantage of the presently disclosed screening method is its continued capacity to select for better and better sesquiterpene synthase variants in an iterative fashion, wherein a test sesquiterpene synthase identified in an iteration is used as the control sesquiterpene synthase in a subsequent iteration. Thus, this method can be distinguished from other assays known in the art that aim to identify only whether a particular sesquiterpene synthase may be active in a biosynthetic pathway, and do not seek to identify synthases having improved activity over a control, *e.g.,* parent synthase. In some embodiments, the FPP level in the host cell is checked and potentially increased at each iteration (*e.g.,* by increasing or decreasing expression levels of enzymes, adding or subtracting enzymes, increasing or decreasing copy numbers of genes, replacing promoters controlling expression of enzymes, or altering enzymes by genetic mutation) to a level that causes reduced viability when the host cell expresses the new control sesquiterpene synthase (*i.e.*, the test sesquiterpene synthase of the previous iteration). Alternatively, or in addition, at each iteration, the expression of the control sesquiterpene

synthase can be reduced (*e.g.*, by decreasing expression of or by using weaker promoters or by reducing the stability of the control sesquiterpene synthase transcript or polypeptide) to provide reduced control sesquiterpene synthase activity. In the next iteration, a test sequiterpene synthase can then be identified that has yet increased *in vivo* performance compared to the test sequiterpene synthase of the previous iteration.

Another major advantage of the presently disclosed screening method is its simplicity and capacity for high-throughput implementation. Sesquiterpene synthase variants that can reduce the intracellular FPP levels in the engineered host cell to non-toxic levels are identified simply based on cell viability, making other costly and time consuming screening methods virtually unnecessary. Thus, in one embodiment, the method is used to screen a collection of sesquiterpene synthase variants (*e.g.*, a library of mutant sesquiterpene synthases) for sesquiterpene synthase variants with improved *in vivo* performance. In such an embodiment, not a single test sesquiterpene synthase is expressed in a host cell but a collection of test sesquiterpene synthases are expressed in a collection of host cells. The host cells can then be grown on agar plates, and host cells expressing sesquiterpene synthase variants with improved *in vivo* performance can be identified based on colony growth. In some embodiments, the collection of sesquiterpene synthase variants comprises from 2 to 5, from 5 to 10, from 10 to 50, from 50 to 100, from 100 to 500, from 500 to 1,000, from 1,000 to 10,000, from 10,000 to 100,000, from 100,000 to 1,000,000, and more, sesquiterpene synthase variants.

Another major advantage of the presently disclosed screening method is that selection for improved sesquiterpene synthases occurs *in vivo* rather than *in vitro*. As a result, improvements of multiple enzyme properties that enhance the *in vivo* performance of the sesquiterpene synthase variant can be obtained.

In another aspect, the present invention provides a second screening method for identifying terpene synthase variants with improved *in vivo* performance. In this second screening method, terpene synthase variants with improved *in vivo* performance are identified by their ability to starve host cells of a polyprenyl diphosphate (*e.g.*, FPP). In the presence of a highly active terpene synthase variant, the intracellular pool of its polyprenyl diphosphate substrate in a host cell may be depleted, causing the cell to not be able to maintain basic cellular processes required for cell survival.

The presently provided second screening method thus comprises the following steps:

a) providing a host cell expressing a control terpene synthase and having a growth rate;

b) expressing in the host cell a test terpene synthase instead of the control terpene synthase, wherein the test terpene synthase is a variant of the control terpene synthase; and

d) identifying the test terpene synthase as having improved *in vivo* performance compared to the control terpene synthase by a decreased growth rate of the host cell expressing the test terpene synthase compared to the host cell expressing the control terpene synthase.

The control terpene synthase may be a monoterpene synthase, a sesquiterpene synthase, a diterpene synthase, a sesterterpene synthase, a triterpene synthase, a tetraterpene synthase, or a polyterpene synthase. In some embodiments, the control terpene synthase is a sesquiterpene synthase. In some such embodiments, the control terpene synthase is a β–farnesene synthase. In some such embodiments, the control terpene synthase is a β–farnesene synthase of *Artemisia annua*. In some such embodiments, the control terpene synthase has an amino acid sequence as given in SEQ ID NO: 111.

The polyprenyl diphosphate substrate that becomes depleted in the host cell in the presence of the test terpene synthase may be FPP. Aside from sesquiterpenes, a number of other compounds are synthesized from FPP that are essential for the viability and growth of the host cell. Such compounds include but are not limited to squalene, lanosterol, ergosterol, cycloartenol, cholesterol, steroid hormones, and vitamin D. Thus, in some embodiments, a host cell expressing the test terpene synthase may comprise reduced amounts of cholesterol or ergosterol in its cell membrane. Methods for the quantification of cholesterol or ergosterol in cells are known in the art (*e.g.*, Crockett and Hazel (2005) J. Experimental Zoology, 271(3): 190-195; Arthington-Skaggs et al. (1999) J Clin Microbiol. 37(10): 3332–3337; Seitz et al. (1979) Physiol. Biochem. 69: 1202-1203). In some embodiments, the basic cellular process required for cell survival that cannot be maintained in the host cell in the presence of the test terpene synthase is the production and/or maintenance of a cell membrane. In other embodiments, the polyprenyl diphosphate substrate that becomes depleted in the host cell in the presence of the test terpene synthase is GPP or GGPP.

In yet another aspect, the present invention provides a competition method for identifying and/or ranking the *in vivo* performance of terpene synthase variants. The competition method employs a known terpene synthase as the comparison enzyme against which the terpene synthase variants are compared. Both the comparison terpene synthase and each of the terpene synthase variants are co-expressed in a host cell in which they then compete for the same polyprenyl diphosphate substrate (*e.g.,* GPP, FPP, or GGPP) to produce their corresponding terpenes. Since the performance of the comparison enzyme remains constant in the host cells, any changes in the ratios of titers of the terpene products produced by the comparison terpene synthase and the terpene synthase variants are the direct result of the activities of the terpene synthase variants. Consequently, such ratios can be used to identify terpene synthase variants with improved *in vivo* performance, and/or to rank or quantitatively compare the terpene synthase variants for their *in vivo* kinetic capacities in diverting polyprenyl diphosphate substrates to the production of terpenes.

The presently provided competition method thus comprises the following steps:

a) dividing a population of host cells into a control population and a test population;

b) expressing in the control population a control terpene synthase and a comparison terpene synthase, wherein the control terpene synthase can convert a polyprenyl diphosphate to a first terpene, and wherein the comparison terpene synthase can convert a polyprenyl diphosphate to a second terpene;

c) expressing in the test population the comparison terpene synthase and a test terpene synthase, wherein the test terpene synthase is a variant of the control terpene synthase, and wherein the comparison terpene synthase is expressed at similar levels in the test population and in the control population; and

d) measuring a ratio of the first terpene over the second terpene in the test population and in the control population.

Notably, the presently disclosed competition method can be applied to a wide variety of terpene synthases. Thus, in separate embodiments, the competition method is applied to identify and/or rank terpene synthases selected from the group consisting of monoterpene synthases, diterpene synthases, sesquiterpene synthases, sesterterpene synthases, triterpene synthases, tetraterpene synthases, and polyterpene synthases. Accordingly, in separate embodiments, the first terpene and the second terpene are selected from the group consisiting of monoterpenes, sesquiterpenes, diterpenes,

sesterterpenes, triterpenes, tetraterpenes, and polyterpenes. In some such embodiments, the first terpene or the second terpene is selected from the group consisting of a β–farnesene, an α–farnesene, a trichodiene, a patchoulol, an amorphadiene, a valencene, a farnesol, a nerolidol, a limonene, a myrcene, and a nootkatone.

The control terpene synthase may be a naturally occurring terpene synthase or a naturally not occurring synthase. The test terpene synthase may comprise amino acid substitutions, deletions, or additions compared to the control terpene synthase, or comprise identical amino acids encoded by different codons in the nucleotide sequences encoding the control terpene synthase and test terpene synthase. In some embodiments, the control terpene synthase is a sesquiterpene synthase. In some such embodiments, the sesquiterpene synthase is selected from the group consisiting of a β–farnesene synthase, an α–farnesene synthase, a trichodiene synthase, a patchoulol synthase, an amorphadiene synthase, a valencene synthase, a farnesol synthase, a nerolidol synthase, and a nootkatone synthase. In some such embodiments, the control sesquiterpene synthase is a β–farnesene synthase of *Artemisia annua*. In some such embodiments, the control sesquiterpene synthase has an amino acid sequence as given in SEQ ID NO: 111.

To be able to compare the ratios of first terpene/second terpene of the control population and the test population, it is necessary to ensure similar expression levels of the comparison terpene synthase. This can be accomplished by placing the nucleotide sequences encoding the comparison terpene synthase in the two host cell populations under the control of the same regulatory elements. In embodiments in which the competition method is used to identify a terpene synthase variant, the expression levels of the control terpene synthase and the test terpene synthase in the two cell populations must also be similar. In other embodiments in which the competition method is used, for example, to identify regulatory elements (*e.g.*, promoters) that provide a desired expression level, the test terpene synthase differs from the control terpene synthase not in nucleotide or amino acid sequence but in expression level. In such embodiments, different regulatory elements are used for the expression of the control terpene synthase and the test terpene synthase.

There are numerous utilities for the presently disclosed competition method. In some embodiments, the method is used to screen for terpene synthase variants with improved *in vivo* performance (*e.g.*, from a library of mutant terpene synthases) on the basis that compared to the control terpene synthase, a terpene synthase variant with

improved *in vivo* performance is capable of diverting more flux from a polyprenyl diphosphate substrate to its terpene product, thus, giving a higher ratio of terpene of interest/comparison terpene (*i.e.*, first terpene/second terpene). In such embodiments, it is important that the test terpene synthase is expressed at a similar level in the test population as the control terpene synthase is expressed in the control population.

A similar assay can be used to rank the strength of a series of promoters (see Example 16, for example, in which such an assay was used to identify promoters suitable for use in expressing the control sesquiterpene synthase in the first screening method disclosed herein). In such an embodiment, the control terpene synthase and the test terpene synthase are actually identical, but they are under regulatory control of different prometers such that the control population and the test population do not differ in the type of test terpene synthase they comprise but in the level of expression of the test terpene synthase. In such an embodiment, comparing the ratio of the first terpene over the second terpene in the test population and in the control population provides information not about the activity of the test terpene synthase but about the strength of the promoter driving the expression of the test terpene synthase.

In addition, this system can be used to modulate the ratio of two or more terpene products made by various cells so that the combined mixture of the various cells with a defined ratio possesses the desired properties of a commercially useful product.

Major advantages of the presently disclosed competition method are that it eliminates cell-to-cell variations in enzyme expression and activity, that it is robust, and that it can be used even when the overall pathway flux to the polyprenyl diphosphate substrate is limiting in the host cell. The latter is important because assays that are based on absolute terpene titer measurements may mask improvements in enzyme activities when terpene titers are capped by the overall pathway flux to the polyprenyl disphosphate substrate.

Enzymes developed using the presently disclosed screening method and/or competition method can be subjected to additional means of optional screening including, but not limited to, a fluorescent screen and/or a direct quantitation of terpene product by gas chromatography. More specifically, this includes a Nile Red-based high throughput fluorescent assay for measuring production of a sesquiterpene such as farnesene, and a gas chromatography (GC)-based direct quantitation method for measuring the titer of a sesquiterpene such as farnesene. The improved enzymes can also be further improved by genetic engineering methods such as induced mutations and the like. As a result,

improvements of multiple enzyme properties that enhance the final enzyme performance are successively accomplished, and the most effective enzyme variants are identified.

The present disclosure also pertains to superior farnesene synthase variants, and host cells comprising such farnesene synthase variants. The farnesene synthase variants were developed using the methods disclosed herein, and show more than a 200% improvement in *in vivo* performance. The farnesene synthase variants have improved catalytic efficiency, *i.e.,* they are able to catalyze their reaction at a faster rate. As such, they are more suitable for commercial production of sesquiterpene products such as farnesene where high yield production is of major importance.

Thus, in yet another aspect, provided herein are isolated β–farnesene synthase variants, and isolated nucleic acids comprising a nucleotide sequence encoding such β–farnesene synthase variants, having an amino acid sequence as given in SEQ ID NO: 111 but comprising one or more amino acid substitutions at positions selected from the group consisting of positions 2, 3, 4, 6, 9, 11, 18, 20, 24, 35, 38, 50, 61, 72, 80, 89, 105, 115, 144, 196, 211, 251, 280, 288, 319, 348, 357, 359, 369, 371, 385, 398, 423, 433, 434, 442, 444, 446, 460, 467, 488, 495, 505, 526, 531, 556, 572, and 575 of SEQ ID NO: 111.

In yet another aspect, the present invention provides a genetically modified host cell that comprises:

(a) a heterologous β–farnesene synthase, wherein the heterologous β–farnesene synthase is a variant of a β–farnesene synthase encoded by SEQ ID NO: 111; and

(b) a MEV pathway or DXP pathway enzyme;

wherein the host cell makes at least 15% more of a β–farnesene compared to a parent cell that comprises the MEV pathway or DXP pathway enzyme and the β–farnesene synthase encoded by SEQ ID NO: 111.

In some embodiments, the heterologous β–farnesene synthase comprises one or more amino acid substitutions at positions selected from the group consisting of positions 2, 3, 4, 6, 9, 11, 18, 20, 24, 35, 38, 50, 61, 72, 80, 89, 105, 115, 144, 196, 211, 251, 280, 288, 319, 348, 357, 359, 369, 371, 385, 398, 423, 433, 434, 442, 444, 446, 460, 467, 488, 495, 505, 526, 531, 556, 572, and 575 of SEQ ID NO: 111.

In some embodiments, the MEV pathway enzyme is a HMG-CoA reductase. In some embodiments, the MEV pathway enzyme is a mevalonate kinase. Additional exemplary enzymes of the MEV pathway are provided in Section 5.4 below.

In yet another aspect, provided herein is a method of producing a β–farnesene comprising the steps of:

(a) obtaining a plurality of genetically modified host cells comprising:

i) a first heterologous nucleotide sequence encoding a variant of a β–farnesene synthase encoded by SEQ ID NO: 111; and

ii) a second heterologous nucleotide sequence encoding a MEV pathway or DXP pathway enzyme;

(b) culturing said genetically modified host cells in a medium comprising a carbon source under conditions suitable for making the β–farnesene; and

(c) recovering the β–farnesene from the medium.

In some embodiments, the MEV pathway enzyme is a HMG-CoA reductase. In some embodiments, the MEV pathway enzyme is a mevalonate kinase. Additional exemplary enzymes of the MEV pathway are provided in Section 5.4 below.

5.3. Selecting Host Cells

Host cells useful in the practice of the present invention include archae, prokaryotic, or eukaryotic cells.

Suitable prokaryotic hosts include but are not limited to any of a variety of gram-positive, gram-negative, or gram-variable bacteria. Examples include but are not limited to cells belonging to the genera: *Agrobacterium, Alicyclobacillus, Anabaena, Anacystis, Arthrobacter, Azobacter, Bacillus, Brevibacterium, Chromatium, Clostridium, Corynebacterium, Enterobacter, Erwinia, Escherichia, Lactobacillus, Lactococcus, Mesorhizobium, Methylobacterium, Microbacterium, Phormidium, Pseudomonas, Rhodobacter, Rhodopseudomonas, Rhodospirillum, Rhodococcus, Salmonella, Scenedesmun, Serratia, Shigella, Staphlococcus, Strepromyces, Synnecoccus, and Zymomonas. Examples of prokaryotic strains include but are not limited to: Bacillus subtilis, Bacillus amyloliquefacines, Brevibacterium ammoniagenes, Brevibacterium immariophilum, Clostridium beigerinckii, Enterobacter sakazakii, Escherichia coli, Lactococcus lactis, Mesorhizobium loti, Pseudomonas aeruginosa, Pseudomonas mevalonii, Pseudomonas pudica, Rhodobacter capsulatus, Rhodobacter sphaeroides, Rhodospirillum rubrum, Salmonella enterica, Salmonella typhi, Salmonella typhimurium, Shigella dysenteriae, Shigella flexneri, Shigella sonnei,* and *Staphylococcus aureus*.

Suitable archae hosts include but are not limited to cells belonging to the genera: *Aeropyrum, Archaeglobus, Halobacterium, Methanococcus, Methanobacterium,*

Pyrococcus, Sulfolobus, and *Thermoplasma*. Examples of archae strains include but are not limited to: *Archaeoglobus fulgidus, Halobacterium sp., Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Thermoplasma acidophilum, Thermoplasma volcanium, Pyrococcus horikoshii, Pyrococcus abyssi, and Aeropyrum pernix*.

Suitable eukaryotic hosts include but are not limited to fungal cells, algal cells, insect cells, and plant cells. Examples include but are not limited to cells belonging to the genera: *Aspergillus, Candida, Chrysosporium, Cryotococcus, Fusarium, Kluyveromyces, Neotyphodium, Neurospora, Penicillium, Pichia, Saccharomyces, Trichoderma, Ascomycota, Basidiomycota, Dothideomycetes*, and *Xanthophyllomyces* (formerly *Phaffia*). Examples of eukaryotic strains include but are not limited to: *Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia quercuum, Pichia pijperi, Pichia stiptis, Pichia methanolica, Pichia sp., Saccharomyces cerevisiae, Saccharomyces sp., Schizosaccharomyces pombe, Hansenula polymorpha, Kluyveromyces sp. , Kluyveromyces lactis, Candida albicans, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Trichoderma reesei, Chrysosporium lucknowense, Fusarium sp., Fusarium gramineum, Fusarium venenatum, Neurospora crassa*, and *Chlamydomonas reinhardtii*.

In a particular embodiment, the host cell is an *Escherichia coli* cell. In another particular embodiment, the host cell is a *Saccharomyces cerevisiae* cell. In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell selected from the group consisting of Baker's yeast, CBS 7959, CBS 7960, CBS 7961, CBS 7962, CBS 7963, CBS 7964, IZ-1904, TA, BG-1, CR-1, SA-1, M-26, Y-904, PE-2, PE-5, VR-1, BR-1, BR-2, ME-2, VR-2, MA-3, MA-4, CAT-1, CB-1, NR-1, BT-1, and AL-1. In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell selected from the group consisting of PE-2, CAT-1, VR-1, BG-1, CR-1, and SA-1. In a particular embodiment, the host cell is a *Saccharomyces cerevisiae* of strain PE-2. In another particular embodiment, the host cell is a *Saccharomyces cerevisiae* of strain CAT-1. In another particular embodiment, the host cell is a *Saccharomyces cerevisiae* of strain BG-1.

In some embodiments, the host cell is a cell that is suitable for industrial fermentation, *e.g.,* bioethanol fermentation. In particular embodiments, the host cell is conditioned to subsist under high solvent concentration, high temperature, expanded substrate utilization, nutrient limitation, osmotic stress, acidity, sulfite and bacterial

contamination, or combinations thereof, which are recognized stress conditions of the industrial fermentation environment.

5.4. Host Cells With Elevated Intracellular FPP Levels

In some embodiments, compared to a parent cell, a host cell comprises an elevated intracellular level of FPP, wherein the elevated intracellular level of FPP decreases the viability of the host cell.

In some embodiments, the host cell comprises an intracellular level of FPP that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the intracellular level of FPP of the parent cell, on a per unit volume of cell culture basis.

In some embodiments, the host cell comprises an intracellular level of FPP that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the intracellular level of FPP of the parent cell, on a per unit dry cell weight basis.

In some embodiments, the host cell comprises an intracellular level of FPP that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher

than the intracellular level of FPP of the parent cell, on a per unit volume of cell culture per unit time basis.

In some embodiments, the host cell comprises an intracellular level of FPP that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the intracellular level of FPP of the parent cell, on a per unit dry cell weight per unit time basis.

In most embodiments, the elevated intracellular level of FPP in the host cell is inducible by an inducing compound. Such a host cell can be manipulated with ease in the absence of the inducing compound. The inducing compound is then added to induce the elevated level of FPP in the host cell. In other embodiments, the elevated intracellular level of FPP in the host cell is inducible by changing culture conditions, such as, for example, the growth temperature. The inducible elevation of intracellular FPP level thus provides a molecular on and off switch for the reduced viability phenotype of the host cell.

The elevation of intracellular FPP level can be effected through targeted genetic engineering of the host cell. A number of enzymes are known to function in the production or utilization of FPP and its precursors, and any one of these enzymes can be manipulated to change the level of FPP in a host cell.

In some embodiments, the production of FPP in the host cell is increased by increasing production of cellular acetyl-CoA in the host cell.

In some embodiments, the production of FPP in the host cell is increased by increasing the production of IPP and/or DMAPP in the host cell. In some such embodiments, the production of IPP and DMAPP in the host cell is increased by increasing the activity of one or more enzymes of the MEV pathway. A schematic representation of the MEV pathway is described in FIG. 15. In general, the pathway comprises six steps:

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. 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*).

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*).

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

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*).

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*).

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*).

In other such embodiments, the production of IPP and DMAPP in the host cell is increased by increasing the activity of one or more enzymes of the DXP pathway. A schematic representation of the DXP pathway is described in FIG. 16. 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; *Rhodopseudomonas palustris* CGA009), (NC_004556, locus tag PD1293; *Xylella fastidiosa Temecula1*), and (NC_003076, locus tag AT5G11380; *Arabidopsis thaliana*).

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).

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).

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).

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).

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).

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).

In some embodiments, the production of FPP in the host cell is increased by increasing the isomerization of IPP to DMAPP. In some such embodiments, the isomerization of IPP to DMAPP is increased by increasing the activity of an IPP isomerase. Illustrative examples of nucleotide sequences encoding IPP isomerases include but are not limited to: (NC_000913, 3031087. . 3031635; *Escherichia coli*), and (AF082326; *Haematococcus pluvialis*).

In some embodiments, the production of FPP in the host cell is increased by increasing the condensation of IPP and DMAPP to FPP. In some such embodiments, the condensation of IPP and DMAPP or of IPP and geranyl pyrophosphate ("GPP") to FPP is increased by increasing the activity of a FPP synthase. Illustrative examples of nucleotide sequences that encode FPP synthases include but are not limited to: (ATU80605; *Arabidopsis thaliana*), (ATHFPS2R; *Arabidopsis thaliana*), (AAU36376; *Artemisia annua*), (AF461050; *Bos taurus*), (D00694; *Escherichia coli* K-12), (AE009951, Locus AAL95523; *Fusobacterium nucleatum subsp. nucleatum* ATCC 25586), (GFFPPSGEN; *Gibberella fujikuroi*), (CP000009, Locus AAW60034; *Gluconobacter oxydans* 621H), (AF019892; *Helianthus annuus*), (HUMFAPS; *Homo sapiens*), (KLPFPSQCR; *Kluyveromyces lactis*), (LAU15777; *Lupinus albus*), (LAU20771; *Lupinus albus*),

(AF309508; *Mus musculus*), (NCFPPSGEN; *Neurospora crassa*), (PAFPS1; *Parthenium argentatum*), (PAFPS2; *Parthenium argentatum*), (RATFAPS; *Rattus norvegicus*), (YSCFPP; *Saccharomyces cerevisiae*), (D89104; *Schizosaccharomyces pombe*), (CP000003, Locus AAT87386; *Streptococcus pyogenes*), (CP000017, Locus AAZ51849; *Streptococcus pyogenes*), (NC_008022, Locus YP_598856; *Streptococcus pyogenes* MGAS10270), (NC_008023, Locus YP_600845; *Streptococcus pyogenes* MGAS2096), (NC_008024, Locus YP_602832; *Streptococcus pyogenes* MGAS10750), (MZEFPS; *Zea mays*), (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 cerevisae*), (CP000031, Locus AAV93568; *Silicibacter pomeroyi* DSS-3), (AE008481, Locus AAK99890; *Streptococcus pneumoniae* R6), and (NC_004556, Locus NP 779706; *Xylella fastidiosa* Temecula1).

In some embodiments, the production of FPP in the host cell is increased by inhibiting reactions that divert intermediates from productive steps towards formation of FPP. Such reactions include but are not limited to side reactions of the TCA cycle that lead to fatty acid biosynthesis, alanine biosynthesis, the aspartate superpathway, gluconeogenesis, heme biosynthesis, glutamate biosynthesis, and conversion of acetyl-CoA to acetate via the action of phosphotransacetylase.

In some embodiments, a host cell that comprises an elevated intracellular level of FPP is obtained by decreasing the consumption of FPP in the host cell. In some such embodiments, the consumption of FPP in the host cell is decreased by decreasing the activity of a farnesyl-diphosphate farnesyl transferase or squalene synthase that can convert FPP to squalene. In other such embodiments, the consumption of FPP in the host cell is decreased by decreasing the activity of a sesquiterpene synthase in the host cell.

A host cell comprising an elevated intracellular level of FPP can be generated by genetically modifying a parent cell using genetic engineering techniques (*i.e.*, recombinant technology), classical microbiological techniques, or a combination of such techniques.

The host cell may also be a naturally occurring genetic variant that is non-viable under certain growth conditions due to an elevated intracellular level of FPP.

A host cell that comprises such an elevated intracellular level of FPP that it has reduced cell viability can be identified by comparing the growth of the host cell on a solid medium with that of a parent cell that does not comprise the elevated intracellular level of FPP. A host cell that comprises an elevated level of intracellular FPP should produce fewer or smaller colonies on the solid agar medium compared to its parent cell. A host cell that comprises the elevated intracellular level of FPP only under certain growth conditions can be identified by first growing the host cell under conditions under which the host cell does not comprise an elevated intracellular level of FPP and under which it has the same viability as its parent cell ("permissive growth conditions"), and then replica-plating the host cell and growing it under conditions under which the host cell does comprise the elevated intracellular level of FPP ("restrictive growth condition") to identify host cells that has reduced viability only under restrictive growth conditions but that does not have reduced viability under permissive growth conditions. Such restrictive growth conditions can include but are not limited to the presence of a specific nutrient in the culture medium, the presence of a specific nutrient at a specific level in the culture medium, the presence of an inducing compound in the culture medium, the presence of a repressing compound in the culture medium, and a specific growth temperature.

5.5. Terpene Synthases

The methods provided herein are focused on developing terpene synthase variants with improved *in vivo* performance.

In some embodiments, the terpene synthase variant is a variant of a naturally occurring terpene synthase. In other embodiments, the terpene synthase variant is a variant of a naturally not occurring terpene synthase.

In some such embodiments, the terpene synthase variant differs from a naturally occurring terpene synthase or from a naturally not occurring terpene synthase by one or more amino acid substitutions, deletions, and/or additions. In some embodiments, the terpene synthase differs from a naturally occurring terpene synthase or from a naturally not occurring terpene synthase by comprising one, two, three, four, five, six, seven, eight, nine, ten, or more additional amino acids. In some embodiments, the terpene synthase variant differs from a naturally occurring terpene synthase or from a naturally not occurring terpene synthase by comprising one, two, three, four, five, six, seven, eight, nine, ten, or more amino acid substitutions. In some embodiments, the terpene synthase

variant differs from a naturally occurring terpene synthase or from a naturally not occurring terpene synthase by lacking one, two, three, four, five, six, seven, eight, nine, ten, or more amino acids.

In some embodiments, the terpene synthase variant has from about 50% to about 55%, from about 55% to about 60%, from about 60% to about 65%, from about 65% to about 70%, from about 70% to about 75%, from about 75% to about 80%, from about 80% to about 85%, from about 85% to about 90%, from about 90% to about 95%, or from about 95% to 99% amino acid sequence identity to the amino acid sequence of a naturally occurring terpene synthase or of a naturally not occurring terpene synthase.

In some embodiments, the terpene synthase variant comprises a consensus amino acid sequence. A consensus amino acid sequence is derived by aligning three or more amino acid sequences, and identifying amino acids that are shared by at least two of the sequences. In some embodiments, the terpene synthase variant comprises a consensus sequence derived from two or more naturally occurring terpene synthases.

In some embodiments, the terpene synthase variant is a hybrid terpene synthase. Hybrid terpene synthases comprise stretches of contiguous amino acids from two or more different terpene synthases. Hybrid terpene synthases can be generated using any known method, including but not limited to exon shuffling, domain swapping, and the like (*e.g.*, Nixon *et al.* (1997) *Proc. Natl. Acad. Sci.* USA 94:1069-1073; Fisch *et al.* (1996) *Proc Natl Acad Sci* USA 93(15):7761-7766).

In some embodiments, a nucleic acid comprising a nucleotide sequence encoding a terpene synthase variant hybridizes under stringent hybridization conditions to a nucleic acid encoding a naturally occurring terpene synthase. In another embodiment, a nucleic acid comprising a nucleotide sequence encoding a terpene synthase variant hybridizes under moderate hybridization conditions to a nucleic acid encoding a naturally occurring terpene synthase. In yet another embodiment, a nucleic acid comprising a nucleotide sequence encoding a terpene synthase variant hybridizes under low stringency hybridization conditions to a nucleic acid encoding a naturally occurring terpene synthase.

In some embodiments, the nucleotide sequence encoding the terpene synthase variant is altered from the nucleotide sequence encoding a naturally occurring terpene synthase to reflect the codon preferences for a particular host cell (*i.e.*, is codon-optimized for expression in a particular host cell). The use of preferred codons for a particular host cell generally increases the likelihood of translation, and hence expression, of the nucleotide sequence. Codon usage tables that summarize the percentage of time a specific

organism uses a specific codon to code a specific amino acid are available for many organisms, and can be used as a reference in designing suitable nucleotide sequences. In some embodiments, the nucleotide sequence encoding the terpene synthase is altered to reflect the codon preferences of *Saccharomyces cerevisiae* (see, *e.g.*, Bennetzen and Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031). In some embodiments, the nucleotide sequence encoding the terpene synthase is altered to reflect the codon preferences for *Escherichia coli* (see, *e.g.*, Gouy and Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13(6):864-872; Nakamura *et al.* (2000) *Nucleic Acids Res.* 28(1):292).

A nucleic acid comprising a nucleotide sequence encoding a terpene synthase can be obtained using any of a variety of known recombinant techniques and synthetic procedures. The nucleic acid can be prepared from genomic DNA, cDNA, or RNA, all of which can be extracted directly from a cell or can be recombinantly produced by various amplification processes including but not limited to PCR and rt-PCR. Direct chemical synthesis methods are also well known in the art.

A nucleic acid comprising a nucleotide sequence encoding a terpene synthase variant can be obtained using any of a variety of known methods. For example, nucleic acids can be isolated from cells that were treated with chemical mutagens or radiation, or from cells that have deficiencies in DNA repair. Suitable chemical mutagens include, but are not limited to, ethyl methanesulfonate (EMS), methyl methanesulfonate (MMS), N-nitroso urea (ENU), N-methyl-N-nitro-N'-nitrosoguanidine, 4-nitroquinoline N-oxide, diethylsulfate, benzopyrene, cyclophosphamide, bleomycin, triethylmelamine, acrylamide monomer, nitrogen mustard, vincristine, diepoxyalkanes (for example, diepoxybutane), ICR-170, formaldehyde, procarbazine hydrochloride, ethylene oxide, dimethylnitrosamine, 7,12 dimethylbenz(a)anthracene, chlorambucil, hexamethylphosphoramide, bisulfan, and acridine dyes (see, for example Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology, Second Edition* (1989) Sinauer Associates, Inc. , Sunderland, Mass., or Deshpande Mukund V., *Appl. Biochem. Biotechnol.* 36, 227 (1992)). Suitable radiation exposures include but are not limited to ultraviolet radiation (optionally in combination with exposure to chemical agents such as, for example, trimethylpsoralen), γ-irradiation, X-rays, and fast neutron bombardment. A suitable method for introducing deficiencies in DNA repair in a cell includes but is not limited to the expression of a mutant DNA repair enzyme that generates a high frequency of mutations in the genome of the cell (on the order of about 1 mutation/100 genes to

about 1 mutation/10,000 genes). Examples of genes encoding DNA repair enzymes include but are not limited to Mut H, Mut S, Mut L, and Mut U, and the homologs thereof in other species (for example, MSH 1-6, PMS 1-2, MLH 1, GTBP, and ERCC-1). Other methods for obtaining a nucleic acid comprising a nucleotide sequence encoding a terpene synthase variant include manipulation of cell-free *in vitro* systems (*e.g.*, using error-prone PCR for the amplification of a nucleic acid), random or targeted insertion in the genome of a cell of a mobile DNA element (*e.g.*, a transposable element), or *in vitro* DNA shuffling (*e.g.*, exon shuffling, domain swapping, and the like; see, for example, Ausubel *et al.*, *Current Protocols In Molecular Biology*, John Wiley and Sons, New York (current edition); and Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 3d. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001)).

In some embodiments, the terpene synthase variants are variants of a sesquiterpene synthase selected from the group consisiting of a β–farnesene synthase, an α–farnesene synthase, a trichodiene synthase, a patchoulol synthase, an amorphadiene synthase, a valencene synthase, a farnesol synthase, a nerolidol synthase, and a nootkatone synthase.

In some embodiments, the terpene synthase variant is a β–farnesene synthase variant. In some such embodiments, the β–farnesene synthase variant is derived from a β–farnesene synthase of *Artemisia annua*. The sequence of the β–farnesene synthase of *Artemisia annua* has been previously described (Picaud, et al, (2005) *Phytochemistry* 66 (9):961-967). The nucleotide sequence of the β–farnesene synthase of *Artemisia annua* is deposited under GenBank accession number AY835398, and SEQ ID NO: 112 as provided herein. The amino acid sequence of the β–farnesene synthase of *Artemisia annua* is deposited under GenBank accession number AAX39387, and SEQ ID NO: 111 as provided herein.

In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 2 from serine to aspartate (S2D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 3 from threonine to asparagine (T3N mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 4 from leucine to serine (L4S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid

substitution at position 6 from isoleucine to threonine (I6T mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 9 from valine to aspartic acid (V9D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 11 from phenylalanine to serine (F11S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 20 from valine to glutamic acid (V20E mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 24 from valine to aspartic acid (V24D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 35 from methionine to threonine (M35T mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 38 from asparagine to serine (N38S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 50 from aspartic acid to asparagine (D50N mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 61 from leucine to glutamine (L61Q mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 72 from glutamic acid to lysine (E72K mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 72 from glutamic acid to valine (E72V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 80 from asparagine to aspartic acid (N80D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 89 from isoleucine to valine (I89V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 105 from glutamic

acid to aspartic acid (E105D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 115 from isoleucine to methionine (I115M mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 115 from isoleucine to valine (I115V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 144 from phenylalanine to tyrosine (F144Y mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 196 from threonine to serine (T196S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 211 from serine to threonine (S211T mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 251 from leucine to methionine (L251M mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 280 from leucine to glutamine (L280Q mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 288 from tyrosine to phenylalanine (Y288F mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 319 from threonine to serine (T319S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 357 from glutamic acid to valine (E357V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 359 from glutamic acid to threonine (E359T mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 369 from valine to leucine (V369L mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 371 from leucine to methionine (L371M mutation). In some

embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 385 from threonine to alanine (T385A mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 398 from isoleucine to valine (I398V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 423 from valine to isoleucine (V423I mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 433 from methionine to isoleucine (M433I mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 434 from isoleucine to threonine (I434T mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 442 from glycine to alanine (G442A mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 442 from glycine to aspartic acid (G442D mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 444 from isoleucine to leucine (I444L mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 446 from threonine to asparagine (T446N mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 460 from isoleucine to valine (I460V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 467 from valine to isoleucine (V467I mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 488 from serine to phenylalanine (S488F mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 495 from glutamic acid to glycine (E495G mutation). In some embodiments, the β–farnesene synthase variant

has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 505 from glutamic acid to valine (E505V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 526 from threonine to serine (T526S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 531 from proline to serine (P531S mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 556 from alanine to valine (A556V mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 572 from methionine to lysine (M572K mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 575 from a stop codon to lysine (stop575K mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 348 from arginine to lysine (R348K mutation). In some embodiments, the β–farnesene synthase variant has an amino acid sequence as given in SEQ ID NO: 111 but comprising an amino acid substitution at position 18 from leucine to isoleucine (L18I mutation).

5.6. Genetically Engineering Host Cells

The methods provided herein include obtaining a host cell that is genetically engineered to comprise an elevated intracellular FPP level or to express a terpene synthase or a terpene synthase variant. Such a genetically engineered host cell may comprise insertions, deletions, or modifications of nucleotides in such a manner as to provide the desired effect of elevating the intracellular level of FPP or of expressing the terpene synthase or the terpene synthase variant. Such genetic modifications may result in a decrease or increase or modification in copy number or activity of a specific enzyme.

For example, the copy number of an enzyme in a host cell may be altered by modifying the transcription of the gene that encodes the enzyme. This can be achieved for example by modifying the copy number of the nucleotide sequence encoding the enzyme (*e.g.*, by using a higher or lower copy number expression vector comprising the nucleotide sequence, or by introducing additional copies of the nucleotide sequence into the genome

of the host cell or by deleting or disrupting the nucleotide sequence in the genome of the host cell), by changing the order of coding sequences on a polycistronic mRNA of an operon or breaking up an operon into individual genes each with its own control elements, or by increasing the strength of the promoter or operator to which the nucleotide sequence is operably linked. Alternatively or in addition, the copy number of an enzyme in a host cell may be altered by modifying the level of translation of an mRNA that encodes the enzyme. This can be achived for example by modifying 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.

The activity of an enzyme in a host cell can be altered in a number of ways, including, but not limited to, expressing a modified form of the enzyme that exhibits increased or decreased solubility in the host cell, expressing an altered form of the enzyme that lacks a domain through which the activity of the enzyme is inhibited, expressing a modified form of the enzyme that has a higher or lower Kcat or a lower or higher Km for the substrate, or expressing an altered form of the enzyme that is more or less affected by feed-back or feed-forward regulation by another molecule in the pathway.

The methods provided herein further include steps of expressing a terpene synthase or a terpene synthase variant in a host cell that does not naturally express such terpene synthase or terpene synthase variant. Expression of a terpene synthase or a terpene synthase variant in a host cell can be accomplished by introducing into the host cells a nucleic acid comprising a nucleotide sequence encoding the terpene synthase or terpene synthase variant under the control of regulatory elements that permit expression in the host cell. In some embodiments, the nucleic acid is an extrachromosomal plasmid. In other embodiments, the nucleic acid is a chromosomal integration vector that can integrate the nucleotide sequence into the chromosome of the host cell.

In some embodiments, it is essential that expression levels of terpene synthases or terpene synthase variants in two or more host cells are similar. This can be accomplished using nucleic acids comprising nucleotide sequences encoding the terpene synthase or

terpene synthase variant under the control of the same regulatory elements. Such nucleic acids can be used as extrachromosomal expression vectors or to integrate the nucleotide sequences encoding the terpene synthase or the terpene synthase variant and the regulatory elements into the chromosome of the host cell. Comparable expression levels can also be accomplished by targeting nucleic acids comprising nucleotide sequences encoding the terpene synthase or terpene synthase variant to identical locations in the two or more host cells, thus placing the nucleotide sequences under the control of the same endogenous regulatory elements. In addition to the use of similar regulatory elements, comparable expression levels may also depend on similar copy numbers of the nucleotide sequences in the two or more host cells. Copy numbers can be controlled by the use of similar or identical origins of replications in extrachromosomal expression vectors, or by the use of similar types and numbers of chromosomal integration constructs for the integration of the nucleotide sequences into the chromosome of the two or more host cells. A number of additional features of the nucleic acids can affect the expression level of the encoded terpene synthases or terpene synthase variants (e.g., protein or mRNA stability, sequence of the ribosome binding site, distance between the ribosome binding site and start codon, nature of the upstream and downstream sequences, hairpins and other specialized sequences, and codon usage), and all of these can be modified to ensure similar expression levels when required in the provided methods.

Nucleic acids can be introduced into microorganisms by any method known to one of skill in the art without limitation (*see*, for example, Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci.* USA 75:1292-3; Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376-3385; Goeddel *et al.* , eds, 1990, Methods in Enzymology, vol. 185, Academic Press, Inc. , CA; Krieger, 1990, Gene Transfer and Expression -- A Laboratory Manual, Stockton Press, NY; Sambrook *et al.* , 1989, Molecular Cloning -- A Laboratory Manual, Cold Spring Harbor Laboratory, NY; and Ausubel *et al.* , eds. , Current Edition, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, NY). Exemplary techniques include but are not limited to spheroplasting, electroporation, PEG 1000 mediated transformation, and lithium acetate or lithium chloride mediated transformation.

In some embodiments, a nucleic acid used to genetically modify a host cell comprises one or more selectable markers useful for the selection of transformed host cells and for placing selective pressure on the host cell to maintain the foreign DNA.

In some embodiments, the selectable marker is an antibiotic resistance marker. Illustrative examples of antibiotic resistance markers include but are not limited to the *BLA, NAT1, PAT, AUR1-C, PDR4, SMR1, CAT*, mouse dhfr, *HPH, DSDA, KAN*^{R}, and *SH BLE* gene products. The *BLA* gene product from *E. coli* confers resistance to beta-lactam antibiotics (*e.g.,* narrow-spectrum cephalosporins, cephamycins, and carbapenems (ertapenem), cefamandole, and cefoperazone) and to all the anti-gram-negative-bacterium penicillins except temocillin; the *NAT1* gene product from *S. noursei* confers resistance to nourseothricin; the *PAT* gene product from *S. viridochromogenes* Tu94 confers resistance to bialophos; the *AUR1-C* gene product from *Saccharomyces cerevisiae* confers resistance to Auerobasidin A (AbA); the PDR4 gene product confers resistance to cerulenin; the SMR1 gene product confers resistance to sulfometuron methyl; the CAT gene product from Tn9 transposon confers resistance to chloramphenicol; the mouse dhfr gene product confers resistance to methotrexate; the *HPH* gene product of *Klebsiella pneumonia* confers resistance to Hygromycin B; the *DSDA* gene product of *E. coli* allows cells to grow on plates with D-serine as the sole nitrogen source; the *KAN*^{R} gene of the Tn903 transposon confers resistance to G418; and the *SH BLE* gene product from *Streptoalloteichus hindustanus* confers resistance to Zeocin (bleomycin). In some embodiments, thes antibiotic resistance marker is deleted after the genetically modified host cell disclosed herein is isolated.

In some embodiments, the selectable marker rescues an auxotrophy (*e.g.,* a nutritional auxotrophy) in the genetically modified microorganism. In such embodiments, a parent microorganism comprises a functional disruption in one or more gene products that function in an amino acid or nucleotide biosynthetic pathway and that when non-functional rende a parent cell incapable of growing in media without supplementation with one or more nutrients. Such gene products include but are not limited to the *HIS3, LEU2, LYS1, LYS2, MET15*, *TRP1, ADE2,* and *URA3* gene products in yeast. The auxotrophic phenotype can then be rescued by transforming the parent cell with an expression vector or chromosomal integration construct encoding a functional copy of the disrupted gene product, and the genetically modified host cell generated can be selected for based on the loss of the auxotrophic phenotype of the parent cell. Utilization of the *URA3, TRP1*, and *LYS2* genes as selectable markers has a marked advantage because both positive and negative selections are possible. Positive selection is carried out by auxotrophic complementation of the *URA3, TRP1*, and *LYS2* mutations, whereas negative selection is based on specific inhibitors, *i.e.,* 5-fluoro-orotic acid (FOA), 5-fluoroanthranilic acid, and

α–aminoadipic acid (aAA), respectively, that prevent growth of the prototrophic strains but allows growth of the *URA3, TRP1*, and *LYS2* mutants, respectively.

In other embodiments, the selectable marker rescues other non-lethal deficiencies or phenotypes that can be identified by a known selection method.

5.7. Growing Host Cells

The present invention provides methods for developing terpene synthase variants with improved *in vivo* performance, and for producing terpenes. The methods generally involve growing a host cell under suitable conditions in a suitable medium comprising a carbon source.

Suitable conditions and suitable media for growing microorganisms are well known in the art. In some embodiments, the suitable medium is supplemented with one or more additional agents, such as, for example, an inducing compound (*e.g.*, when one or more nucleotide sequences encoding a gene product are under the control of an inducible promoter), a repressing compound (*e.g.*, when one or more nucleotide sequences encoding a gene product are under the control of a repressible promoter), or a selection agent (*e.g.*, an antibiotic to select for microorganisms comprising the genetic modifications).

In some embodiments, the carbon source is a monosaccharide (simple sugar), a disaccharide, a polysaccharide, a non-fermentable carbon source, or one or more combinations thereof. Non-limiting examples of suitable monosaccharides include glucose, galactose, mannose, fructose, ribose, and combinations thereof. Non-limiting examples of suitable disaccharides include sucrose, lactose, maltose, trehalose, cellobiose, and combinations thereof. Non-limiting examples of suitable polysaccharides include starch, glycogen, cellulose, chitin, and combinations thereof. Non-limiting examples of suitable non-fermentable carbon sources include acetate and glycerol.

In one aspect, the present invention provides a method for identifying a terpene synthase with improved in vivo performance based on the growth rate of a host cell comprising a test terpene synthase. The growth rate of a host cell can be determined, for example, by growing the host cell in liquid medium for a defined period of time, then plating all or an aliquot of the culture on an agar plate, and finally scoring the number of colonies that arise on the agar plate. Alternatively, the growth rate of a host cell is determined by measuring the biomass of a culture after a defined period of time. Biomass can be measured by determining the density of the liquid culture, *e.g.* by UV spectrometry, or by quantifying biomass index molecules such as hexoseamine and ergosterol (Frey et

al. (1992) Biol. Fertil. Soils 13: 229-234; Newell (1992) p. 521–561. *In* G. C. Carroll and D. T. Wicklow (ed.), The fungal community: its organization and role in the ecosystem, 2nd ed. Marcel Dekker Inc., New York).

5.8. Producing Terpenes

The present invention provides methods for producing terpenes.

In some embodiments, the terpene is produced in an amount greater than about 10 grams per liter of fermentation medium. In some such embodiments, the terpene is produced in an amount 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 per liter of cell culture.

In some embodiments, the terpene is produced in an amount greater than about 50 milligrams per gram of dry cell weight. In some such embodiments, the terpene is produced in an amount from about 50 to about 1500 milligrams, 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 per gram of dry cell weight.

In some embodiments, the terpene is produced in an amount that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the amount of the terpene produced by a host cell that does not comprise the first heterologous nucleotide sequence, on a per unit volume of cell culture basis.

In some embodiments, the terpene is produced in an amount that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the amount

of the terpene produced by a host cell that does not comprise the first heterologous nucleotide sequence, on a per unit dry cell weight basis.

In some embodiments, the terpene is produced in an amount that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the amount of the terpene produced by a host cell that does not comprise the first heterologous nucleotide sequence, on a per unit volume of cell culture per unit time basis.

In some embodiments, the terpene is produced in an amount that is at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 2-fold, at least about 2. 5-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, at least about 50-fold, at least about 75-fold, at least about 100-fold, at least about 200-fold, at least about 300-fold, at least about 400-fold, at least about 500-fold, or at least about 1,000-fold, or more, higher than the amount of the terpene produced by a host cell that does not comprise the first heterologous nucleotide sequence, on a per unit dry cell weight per unit time basis.

5.9. Extracting and Quantifying Terpenes

The terpene produced by the genetically modified host cell of the invention may be isolated from the fermentation using any suitable separation and purification methods known in the art.

In some embodiments, an organic phase comprising the terpene is separated from the fermentation by centrifugation. In other embodiments, an organic phase comprising the terpene separates from the fermentation spontaneously. In yet other embodiments, an organic phase comprising the terpene is separated from the fermentation by adding a deemulsifier and/or a nucleating agent into the fermentation reaction. Illustrative examples of deemulsifiers include flocculants and coagulants. Illustrative examples of nucleating agents include droplets of the terpene itself and organic solvents such as dodecane, isopropyl myristrate, and methyl oleate.

In some embodiments, the terpene is separated from other products that may be present in the organic phase. In some embodiments, separation is achieved using adsorption, distillation, gas-liquid extraction (stripping), liquid-liquid extraction (solvent extraction), ultrafiltration, and standard chromatographic techniques.

In some embodiments, the terpene is pure, *e.g.,* at least about 40% pure, at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98% pure, or more than 98% pure, where "pure" in the context of an terpene refers to an terpene that is free from other terpenes or contaminants.

Terpene production can be readily quantified using well-known methods known in the art including but are not limited to gas chromatography (GC), gas chromatography-mass spectrometry (GC/MS), nuclear magnetic resonance (NMR), RAMAN spectroscopy, optical absorption (UV/VIS), infrared spectroscopy (IR), high performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC/MS), ion chromatography-mass spectrometry, thin layer chromatography, pulsed amperometric detection, and UV-vis spectrometry.

Terpenes produced by host cells can be recovered using any of a variety of methods including but not limited to chromatography, extraction, solvent extraction, membrane separation, electrodialysis, reverse osmosis, distillation, chemical derivatization, and crystallization.

Additional processing steps to improve terpene quantification or isolation include but are not limited to breaking open the host cells. Suitable methods include but are not limited to vortexing, sonication, and the use of glass beads. Other processing steps can include centrifugation to remove unwanted cell debris from the supernatant.

EXAMPLES

The following specific examples are intended to illustrate the disclosure and should not be construed as limiting the scope of the claims.

Example 1

This example describes methods for making DNA constructs useful in the generation and characterization of terpene synthase variants.

Expression plasmid pAM36-MevT66 was generated by inserting the MevT66 operon into vector pAM36. Vector pAM36 was generated by removing the tet resistance gene from and adding an oligonucleotide cassette containing AscI-SfiI-AsiSI-XhoI-PacI-FsIl-PmeI restriction sites into the pACYC184 vector

(GenBank accession number XO6403). The MevT66 operon encoded 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 MevT66 operon was synthetically generated and comprised the atoB gene of *Escherichia coli* (GenBank accession number NC_000913 REGION: 2324131..2325315; encodes an acetoacetyl-CoA thiolase) codon-optimized for expression in *Escherichia coli*, the coding sequence of the ERG13 gene of *Saccharomyces cerevisiae* (GenBank accession number X96617, REGION: 220..1695; encodes a HMG-CoA synthase) codon-optimized for expression in *Escherichia coli*, and a truncated coding sequence of the HGM1 gene of *Saccharomyces cerevisiae* (GenBank accession number M22002, REGION: 1777..3285; encodes a truncated HMG-CoA reductase) codon-optimized for expression in *Escherichia coli*. The synthetically generated MevT66 operon was cloned into a cloning vector such as a standard pUC or pACYC origin vector, from which it was again PCR amplified with flanking SfiI and AsiSI restriction sites, the amplified DNA fragment was digested using SfiI and AsiSI restriction endonucleases, the approximately 4.2 kb DNA fragment comprising the MevT66 operon was gel purified, and the purified DNA fragment was inserted into the SfiI and AsiSI restriction sites of the pAM36 vector, yielding expression plasmid pAM36-MevT66.

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 coding sequences of the ERG12 gene (GenBank accession number X55875, REGION: 580..1911; encodes a mevalonate kinase), the coding sequence of the ERG8 gene (GenBank accession number Z49939, REGION: 3363..4718; encodes a phosphomevalonate kinase), and the coding sequence of the MVD1 gene (GenBank accession number X97557, REGION: 544..1734; encodes a mevalonate pyrophosphate carboxylase) were PCR amplified from *Saccharomyces cerevisiae* genomic DNA. By choosing appropriate primer sequences, the stop codons of the ERG12 and ERG8 coding sequences were changed from TAA to TAG during PCR amplification to introduce ribosome binding sites. The PCR products were spliced together into the MevB operon by sequence overlap extension (SOE; *Ho, et al, 1989*). 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 again by digesting the cloning construct using PstI restriction endonuclease, the approximately 4.2 kb DNA

fragment comprising the MevB operon was gel purified, and purified DNA fragment was ligated into the PstI restriction site of vector pBBR1MCS-1 (Kovach *et al.*, Gene 166(1): 175-176 (1995)), yielding expression plasmid pMevB-Cm.

Expression plasmid pMBI was generated by inserting the MBI operon into the pBBR1MCS-3 vector. The MBI operon encodes the same enzymes as the MevB operon, as well as an isopentenyl pyrophosphatase isomerase that catalyzes the conversion of IPP to DMAPP. The MBI operon was generated by PCR amplifying from *Escherichia coli* genomic DNA the coding sequence of the idi gene (GenBank accession number AF119715) using primers that contained an XmaI restriction site at their 5' ends. The PCR product was digested using XmaI restriction endonuclease, the 0.5 kb DNA fragment comprising the idi coding sequence was gel purified, and the purified DNA fragment was inserted into the XmaI restriction site of expression plasmid pMevB-Cm, placing idi at the 3' end of the MevB operon. The MBI operon was then subcloned into the SalI and SacI restriction sites of vector pBBR1MCS-3 (Kovach *et al.*, Gene 166(1): 175-176 (1995)), yielding expression plasmid pMBI.

Expression plasmid pMBIS was generated by inserting the ispA gene into expression plasmid pMBI. The ispA gene encodes a farnesyl pyrophosphate synthase that catalyzes the condensation 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 site and a reverse primer with a SacI restriction site. The amplified PCR product was digested using SacII and SacI restriction endonucleases, the 0.9 kb DNA fragment comprising the ispA coding sequence was gel purified, and the purified DNA fragment was ligated into the SacII and SacI restriction sites of pMBI, placing the ispA coding sequence 3' of idi and the MevB operon, and yielding expression plasmid pMBIS.

Expression plasmid pAM25 was generated by inserting the MevT66 operon into the pAM29 vector. The pAM29 vector was created by assembling the p15A origin of replication and kanamycin resistance conferring gene from the pZS24-MCS1 vector (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 description for pAM36-MevT66 above) was digested using EcoRI and Hind III restriction endonucleases, the approximately 4.2 kb DNA fragment comprising the MevT66 operon was gel purified, and the purified DNA fragment was ligated into the EcoRI and HindIII restriction sites of pAM29, yielding expression plasmid pAM25.

Expression plasmid pAM41 was generated by replacing in expression plasmid pAM25 the truncated coding sequence of the HMG1 gene, which encodes a truncated version of 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 *Staphyloccoccus aureus subsp. aureus* (ATCC 70069) genomic DNA using primers comprising SpeI restriction sites, the PCR product was digested using SpeI restriction endonuclease, and the approximately 1.3 kb DNA fragment comprising the mvaA coding sequence was gel purified. Expression plasmid pAM25 was digested using HindIII restriction endonuclease, terminal overhangs were blunted using T4 DNA polymerase, the linear vector backbone was partially digested using SpeI restriction endonuclease, and the approximately 4.8 kb DNA fragment lacking the truncated HMG1 coding sequence was gel purified. The purified DNA fragments were ligated, yielding expression plasmid pAM41.

Expression plasmid pAM43 was generated by inserting the MBIS operon into expression plasmid pAM36-MevT66. The MBIS operon was PCR amplified from pMBIS using primers comprising a 5' XhoI restriction site and a 3' PacI restriction site, the amplified PCR product was digested using XhoI and PacI restriction endonucleases, the approximately 5.4 kb DNA fragment comprising the MBIS operon was gel purified, and the purified DNA fragment was ligated into the XhoI PacI restriction site of expression plasmid pAM36-MevT66, yielding expression plasmid pAM43.

Expression plasmid pAM45 was generated by inserting lacUV5 promoters in front of the MBIS and MevT66 operons of expression plasmid pAM43. A DNA fragment comprising a nucleotide sequence encoding the lacUV5 promoter was synthesized from oligonucleotides, and inserted into the AscI SfiI and AsiSI XhoI restriction sites of pAM43, yielding expression plasmid pAM45.

Expression plasmid pAM52 was generated by replacing in expression plasmid pAM41 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 *Staphyloccoccus aureus subsp. aureus* (ATCC 70069) genomic DNA, and the amplified DNA fragment was used as a PCR primer to replace the coding

sequence of the HMG1 gene in pAM41 according to the method of Geiser *et al.* (BioTechniques 31:88-92 (2001)), yielding expression plasmid pAM52.

Expression plasmid pAM97 was generated by replacing in expression plasmid pAM45 the MevT66 operon with the (atoB(opt):mvaS:mvaA) operon of expression plasmid pAM52. Expression plasmid pAM45 was digested using AsiSI and SfiI restriction endonucleases, and the approximately 8.3 kb DNA fragment lacking the MevT66 operon was gel purified. The (atoB(opt):mvaS:mvaA) operon of pAM52 was PCR amplified using primers comprising a SfiI and AsiSI restriction site, the PCR product was digested using SfiI and AsiSI restriction endonucleases, and the approximately 3.8 kb DNA fragment comprising the (atoB(opt):mvaS:mvaA) operon was gel purified. The purified DNA fragments were ligated, yielding expression plasmid pAM97.

Expression plasmid pAM765 was generated by replacing in expression plasmid pAM97 the coding sequence of the ERG12 gene, which encodes the *Saccharomyces cerevisiae* mevalonate kinase, with the coding sequence of the mvaK1 gene, which encodes the *Staphylococcus aureus* mevalonate kinase (GenBank accession number AAG02424). The *Staphylococcus aureus* mevalonate kinase is less sensitive to feedback inhibition by FPP (Voynova *et al.* (2004) *J. Bacteriol.* 186:61-67), and so expression plasmid pAM765 can cause greater production of FPP in a host cell than expression plasmid pAM97. The coding sequence of the mvaK1 gene was PCR amplified from an expression plasmid, and the approximately 0.9 kb PCR product was gel purified. The PMK-PMD-idi-ispA operon was PCR amplified from pAM97, and the approximately 4.1 kb PCR product was gel purified. The purified PCR products were stitched together, and the stitched product was gel purified. The purified stitched product and pAM97 were digested using XhoI and SacI restriction endonucleases, the digested DNA fragments were gel purified, and the purified DNA fragments were ligated, yielding expression plasmid pAM765 (SEQ ID NO: 1).

Plasmid pAM489 was generated by inserting the P_{GAL10}-ERG20_P_{GAL1}-tHMGR insert of vector pAM471 into vector pAM466. Vector pAM471 was generated by inserting DNA fragment P_{GAL10}-ERG20_P_{GAL1}-tHMGR, which comprises the coding sequence of the ERG20 gene of *Saccharomyces cerevisiae* (ERG20 nucleotide positions 1 to 1208; A of ATG start codon is nucleotide 1) (ERG20), the genomic locus containing the divergent GAL1 and GAL10 promoter of *Saccharomyces cerevisiae* (GAL1 nucleotide position -1 to -668) (P_{GAL}), and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae* (HMG1 nucleotide positions 1586 to 3323) (tHMGR), into the TOPO Zero

Blunt II cloning vector (Invitrogen, Carlsbad, CA). Vector pAM466 was generated by inserting DNA fragment $TRP1^{-856\ to\ +548}$, which comprises a segment of the wild-type TRP1 locus of *Saccharomyces cerevisiae* that extends from nucleotide position -856 to position 548 and harbors a non-native internal *XmaI* restriction site between bases -226 and -225, into the TOPO TA pCR2.1 cloning vector (Invitrogen,q Carlsbad, CA). DNA fragments P_{GAL10}-ERG20_P_{GAL1}-tHMGR and $TRP1^{-856\ to\ +548}$ were generated by PCR amplification as outlined in Table 1. For the construction of pAM489, 400 ng of pAM471 and 100 ng of pAM466 were digested to completion using *XmaI* restriction enzyme (New England Biolabs, Ipswich, MA), DNA fragments corresponding to the P_{GAL10}-ERG20_P_{GAL1}-tHMGR insert and the linearized pAM466 vector were gel purified, and 4 molar equivalents of the purified insert was ligated with 1 molar equivalent of the purified linearized vector, yielding pAM489. FIG. 1R shows a map and SEQ ID NO: 2 the nucleotide sequence of the TRP1_P_{GAL10}-ERG20_P_{GAL1}-tHMGR_TRP insert of pAM489.

Table 1 – PCR amplifications performed to generate pAM489

PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y051 genomic DNA	61-67-CPK001-G (SEQ ID NO: 3)	61-67-CPK002-G (SEQ ID NO: 4)	$TRP1^{-856\ to\ -226}$
		61-67-CPK003-G (SEQ ID NO: 5)	61-67-CPK004-G (SEQ ID NO: 6)	$TRP1^{-225\text{-}to\ +548}$
	100 ng of EG123 genomic DNA	61-67-CPK025-G (SEQ ID NO: 7)	61-67-CPK050-G (SEQ ID NO: 8)	ERG20
	100 ng of Y002 genomic DNA	61-67-CPK051-G (SEQ ID NO: 9)	61-67-CPK052-G (SEQ ID NO: 10)	$P_{GAL1/10}$
		61-67-CPK053-G (SEQ ID NO: 11)	61-67-CPK031-G (SEQ ID NO: 12)	tHMGR
2	100 ng each of $TRP1^{-856\ to\ -226}$ and $TRP1^{-225\text{-}to\ +548}$ purified PCR products	61-67-CPK001-G (SEQ ID NO: 3)	61-67-CPK004-G (SEQ ID NO: 6)	**$TRP1^{-856\ to\ +548}$**
	100 ng each of ERG20 and $P_{GAL1/10}$ purified PCR products	61-67-CPK025-G (SEQ ID NO: 7)	61-67-CPK052-G (SEQ ID NO: 10)	ERG20-$P_{GAL1/10}$
3	100 ng each of ERG20-$P_{GAL1/10}$ and tHMGR purified PCR products	61-67-CPK025-G (SEQ ID NO: 7)	61-67-CPK031-G (SEQ ID NO: 12)	**P_{GAL10}-ERG20_P_{GAL1}-tHMGR**

Plasmid pAM491 was generated by inserting the P_{GAL10}-ERG13_P_{GAL1}-tHMGR insert of vector pAM472 into vector pAM467. Vector pAM472 was generated by inserting DNA fragment P_{GAL10}-ERG13_P_{GAL1}-tHMGR, which comprises the coding sequence of the ERG13 gene of *Saccharomyces cerevisiae* (ERG13 nucleotide positions 1 to 1626) (ERG13), the genomic locus containing the divergent GAL1 and GAL10 promoter of *Saccharomyces cerevisiae* (GAL1 nucleotide position -1 to -668) (P_{GAL}), and a truncated

ORF of the HMG1 gene of *Saccharomyces cerevisiae* (HMG1 nucleotide position 1586 to 3323) (tHMGR), into the TOPO Zero Blunt II cloning vector. Vector pAM467 was generated by inserting DNA fragment URA3$^{-723\ to\ 701}$, which comprises a segment of the wild-type URA3 locus of *Saccharomyces cerevisiae* that extends from nucleotide position -723 to position -224 and harbors a non-native internal *XmaI* restriction site between bases -224 and -223, into the TOPO TA pCR2.1 cloning vector. DNA fragments P_{GAL10}-ERG13_P_{GAL1}-tHMGR and URA3$^{-723\ to\ 701}$ were generated by PCR amplification as outlined in Table 2. For the construction of pAM491, 400 ng of pAM472 and 100 ng of pAM467 were digested to completion using *XmaI* restriction enzyme, DNA fragments corresponding to the ERG13-P_{GAL}-tHMGR insert and the linearized pAM467 vector were gel purified, and 4 molar equivalents of the purified insert was ligated with 1 molar equivalent of the purified linearized vector, yielding pAM491. FIG. 1S shows a map and SEQ ID NO: 13 the nucleotide sequence of the URA3_P_{GAL10}-ERG13_P_{GAL1}-tHMGR_URA3 insert of pAM491.

Table 2 – PCR amplifications performed to generate pAM491

PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y007 genomic DNA	61-67-CPK005-G (SEQ ID NO: 14)	61-67-CPK006-G (SEQ ID NO: 15)	URA3$^{-723\ to\ -224}$
		61-67-CPK007-G (SEQ ID NO: 16)	61-67-CPK008-G (SEQ ID NO: 17)	URA3$^{-223\ to\ 701}$
	100 ng of Y002 genomic DNA	61-67-CPK032-G (SEQ ID NO: 18)	61-67-CPK054-G (SEQ ID NO: 19)	ERG13
		61-67-CPK052-G (SEQ ID NO: 20)	61-67-CPK055-G (SEQ ID NO: 21)	$P_{GAL1/10}$
		61-67-CPK031-G (SEQ ID NO: 22)	61-67-CPK053-G (SEQ ID NO: 23)	tHMGR
2	100 ng each of URA3$^{-723\ to\ -224}$ and URA3$^{-223\ to\ 701}$ purified PCR products	61-67-CPK005-G (SEQ ID NO: 14)	61-67-CPK008-G (SEQ ID NO: 17)	**URA3$^{-723\ to\ 701}$**
	100 ng each of ERG13 and $P_{GAL1/10}$ purified PCR products	61-67-CPK032-G (SEQ ID NO: 18)	61-67-CPK052-G (SEQ ID NO: 20)	ERG13-$P_{GAL1/10}$
3	100 ng each of ERG13-$P_{GAL1/10}$ and tHMGR purified PCR products	61-67-CPK031-G (SEQ ID NO: 22)	61-67-CPK032-G (SEQ ID NO: 18)	**P_{GAL10}-ERG13_P_{GAL1}-tHMGR**

Plasmid pAM493 was generated by inserting the P_{GAL10}-IDI1_P_{GAL1}-tHMGR insert of vector pAM473 into vector pAM468. Vector pAM473 was generated by inserting DNA fragment P_{GAL10}-IDI1_P_{GAL1}-tHMGR, which comprises the coding sequence of the IDI1 gene of *Saccharomyces cerevisiae* (IDI1 nucleotide position 1 to 1017) (IDI1), the genomic locus containing the divergent GAL1 and GAL10 promoter of *Saccharomyces*

cerevisiae (GAL1 nucleotide position -1 to -668) (P_{GAL}), and a truncated ORF of the HMG1 gene of *Saccharomyces cerevisiae* (HMG1 nucleotide positions 1586 to 3323) (tHMGR), into the TOPO Zero Blunt II cloning vector. Vector pAM468 was generated by inserting DNA fragment $ADE1^{-825\ to\ 653}$, which comprises a segment of the wild-type ADE1 locus of *Saccharomyces cerevisiae* that extends from nucleotide position -225 to position 653 and harbors a non-native internal *XmaI* restriction site between bases -226 and -225, into the TOPO TA pCR2.1 cloning vector. DNA fragments P_{GAL10}-IDI1_P_{GAL1}-tHMGR and $ADE1^{-825\ to\ 653}$ were generated by PCR amplification as outlined in Table 3. For the construction of pAM493, 400 ng of pAM473 and 100 ng of pAM468 were digested to completion using *XmaI* restriction enzyme, DNA fragments corresponding to the P_{GAL10}-IDI1_P_{GAL1}-tHMGR insert and the linearized pAM468 vector were gel purified, and 4 molar equivalents of the purified insert was ligated with 1 molar equivalent of the purified linearized vector, yielding vector pAM493. FIG. 1T shows a map and SEQ ID NO: 24 the nucleotide sequence of the ADE1_P_{GAL10}-IDI1_P_{GAL1}-tHMGR_ADE1 insert of pAM493.

Table 3 – PCR amplifications performed to generate pAM493

PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y007 genomic DNA	61-67-CPK009-G (SEQ ID NO: 25)	61-67-CPK010-G (SEQ ID NO: 26)	$ADE1^{-825\ to\ -226}$
		61-67-CPK011-G (SEQ ID NO: 27)	61-67-CPK012-G (SEQ ID NO: 28)	$ADE1^{-225\ to\ 653}$
	100 ng of Y002 genomic DNA	61-67-CPK047-G (SEQ ID NO: 29)	61-67-CPK064-G (SEQ ID NO: 30)	IDI1
		61-67-CPK052-G (SEQ ID NO: 31)	61-67-CPK065-G (SEQ ID NO: 32)	$P_{GAL1/10}$
		61-67-CPK031-G (SEQ ID NO: 12)	61-67-CPK053-G (SEQ ID NO: 33)	tHMGR
2	100 ng each of $ADE1^{-825\ to\ -226}$ and $ADE1^{-225\ to\ 653}$ purified PCR products	61-67-CPK009-G (SEQ ID NO: 25)	61-67-CPK012-G (SEQ ID NO: 28)	**$ADE1^{-825\ to\ 653}$**
	100 ng each of IDI1 and $P_{GAL1/10}$ purified PCR products	61-67-CPK047-G (SEQ ID NO: 29)	61-67-CPK052-G (SEQ ID NO: 31)	IDI1-$P_{GAL1/10}$
3	100 ng each of IDI1-$P_{GAL1/10}$ and tHMGR purified PCR products	61-67-CPK031-G (SEQ ID NO: 12)	61-67-CPK047-G (SEQ ID NO: 29)	**P_{GAL10}-IDI1_P_{GAL1}-tHMGR**

Plasmid pAM495 was generated by inserting the P_{GAL10}-ERG10_P_{GAL1}-ERG12 insert of pAM474 into vector pAM469. Vector pAM474 was generated by inserting DNA fragment P_{GAL10}-ERG10_P_{GAL1}-ERG12, which comprises the coding sequence of the ERG10 gene of *Saccharomyces cerevisiae* (ERG10 nucleotide position 1 to 1347) (ERG10), the genomic locus containing the divergent GAL1 and GAL10 promoter of

Saccharomyces cerevisiae (GAL1 nucleotide position -1 to -668) (P_{GAL}), and the coding sequence of the ERG12 gene of *Saccharomyces cerevisiae* (ERG12 nucleotide position 1 to 1482) (ERG12), into the TOPO Zero Blunt II cloning vector. Vector pAM469 was generated by inserting DNA fragment $HIS3^{-32\ to\ -1000}$_HISMX_$HIS3^{504\ to\ -1103}$, which comprises two segments of the HIS locus of *Saccharomyces cerevisiae* that extend from nucleotide position -32 to position -1000 and from nucleotide position 504 to position 1103, a HISMX marker, and a non-native *XmaI* restriction site between the $HIS3^{504\ to\ -1103}$ sequence and the HISMX marker, into the TOPO TA pCR2.1 cloning vector. DNA fragments P_{GAL10}-ERG10_P_{GAL1}-ERG12 and $HIS3^{-32\ to\ -1000}$_HISMX_$HIS3^{504\ to\ -1103}$ were generated by PCR amplification as outlined in Table 4. For construction of pAM495, 400 ng of pAM474 and 100 ng of pAM469 were digested to completion using *XmaI* restriction enzyme, DNA fragments corresponding to the P_{GAL10}-ERG10_P_{GAL1}-ERG12 insert and the linearized pAM469 vector were gel purified, and 4 molar equivalents of the purified insert was ligated with 1 molar equivalent of the purified linearized vector, yielding vector pAM495. FIG. 1D shows a map and SEQ ID NO: 34 the nucleotide sequence of the HIS3_P_{GAL10}-ERG10_P_{GAL1}-ERG12_HIS3 insert of pAM495.

Table 4 – PCR reactions performed to generate pAM495

PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y007 genomic DNA	61-67-CPK013-G (SEQ ID NO: 35)	61-67-CPK014alt-G (SEQ ID NO: 36)	$HIS3^{-32\ to\ -1000}$
		61-67-CPK017-G (SEQ ID NO: 37)	61-67-CPK018-G (SEQ ID NO: 38)	$HIS3^{504\ to\ -1103}$
		61-67-CPK035-G (SEQ ID NO: 39)	61-67-CPK056-G (SEQ ID NO: 40)	ERG10
		61-67-CPK057-G (SEQ ID NO: 41)	61-67-CPK058-G (SEQ ID NO: 42)	P_{GAL}
		61-67-CPK040-G (SEQ ID NO: 43)	61-67-CPK059-G (SEQ ID NO: 44)	ERG12
	10 ng of plasmid pAM330 DNA **	61-67-CPK015alt-G (SEQ ID NO: 45)	61-67-CPK016-G (SEQ ID NO: 46)	HISMX
2	100 ng each of $HIS3^{504\ to\ -1103}$ and HISMX PCR purified products	61-67-CPK015alt-G (SEQ ID NO: 45)	61-67-CPK018-G (SEQ ID NO: 38)	HISMX_$HIS3^{504\ to\ -1103}$
	100 ng each of ERG10 and P_{GAL} purified PCR products	61-67-CPK035-G (SEQ ID NO: 39)	61-67-CPK058-G (SEQ ID NO: 42)	ERG10-$P_{GAL1/10}$
3	100ng each of $HIS3^{-32\ to\ -1000}$ and HISMX_$HIS3^{504\ to\ -1103}$ purified PCR products	61-67-CPK013-G (SEQ ID NO: 35)	61-67-CPK018-G (SEQ ID NO: 38)	**$HIS3^{-32\ to\ -1000}$_HISMX_$HIS3^{504\ to\ -1103}$**
	100 ng each of ERG10-$P_{GAL1/10}$ and ERG12 purified PCR products	61-67-CPK035-G (SEQ ID NO: 39)	61-67-CPK040-G (SEQ ID NO: 43)	**P_{GAL10}-ERG10_P_{GAL1}-ERG12**

** The HISMX marker in pAM330 originated from pFA6a-HISMX6-PGAL1 as described by van Dijken *et al.* ((2000) Enzyme Microb. Technol. 26(9-10):706-714).

Plasmid pAM497 was generated by inserting the P_{GAL10}-ERG8_P_{GAL1}-ERG19 insert of pAM475 into vector pAM470. Vector pAM475 was generated by inserting DNA fragment P_{GAL10}-ERG8_P_{GAL1}-ERG19, which comprises the coding sequence of the ERG8 gene of *Saccharomyces cerevisiae* (ERG8 nucleotide position 1 to 1512) (ERG8), the genomic locus containing the divergent GAL1 and GAL10 promoter of *Saccharomyces cerevisiae* (GAL1 nucleotide position -1 to -668) (P_{GAL}), and the coding sequence of the ERG19 gene of *Saccharomyces cerevisiae* (ERG19 nucleotide position 1 to 1341) (ERG19), into the TOPO Zero Blunt II cloning vector. Vector pAM470 was generated by inserting DNA fragment $LEU2^{-100\ to\ 450}$_HISMX_$LEU2^{1096\ to\ 1770}$, which comprises two segments of the LEU2 locus of *Saccharomyces cerevisiae* that extend from nucleotide position -100 to position 450 and from nucleotide position 1096 to position 1770, a HISMX marker, and a non-native *XmaI* restriction site between the $LEU2^{1096\ to\ 1770}$ sequence and the HISMX marker, into the TOPO TA pCR2.1 cloning vector. DNA fragments P_{GAL10}-ERG8_P_{GAL1}-ERG19 and $LEU2^{-100\ to\ 450}$_HISMX_$LEU2^{1096\ to\ 1770}$ were generated by PCR amplification as outlined in Table 5. For the construction of pAM497, 400 ng of pAM475 and 100 ng of pAM470 were digested to completion using *XmaI* restriction enzyme, DNA fragments corresponding to the ERG8-P_{GAL}-ERG19 insert and the linearized pAM470 vector were purified, and 4 molar equivalents of the purified insert was ligated with 1 molar equivalent of the purified linearized vector, yielding vector pAM497. FIG. 1V for a map and SEQ ID NO: 47 the nucleotide sequence of the LEU2_P_{GAL10}-ERG8_P_{GAL1}-ERG19_LEU2 insert of pAM497.

Table 5 – PCR reactions performed to generate pAM497				
PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y007 genomic DNA	61-67-CPK019-G (SEQ ID NO: 48)	61-67-CPK020-G (SEQ ID NO: 49)	$LEU2^{-100\ to\ 450}$
		61-67-CPK023-G (SEQ ID NO: 50)	61-67-CPK024-G (SEQ ID NO: 51)	$LEU2^{1096\ to\ 1770}$
	10ng of plasmid pAM330 DNA **	61-67-CPK021-G (SEQ ID NO: 52)	61-67-CPK022-G (SEQ ID NO: 53)	HISMX
	100 ng of Y002 genomic DNA	61-67-CPK041-G (SEQ ID NO: 54)	61-67-CPK060-G (SEQ ID NO: 55)	ERG8
		61-67-CPK061-G (SEQ ID NO: 56)	61-67-CPK062-G (SEQ ID NO: 57)	P_{GAL}
		61-67-CPK046-G (SEQ ID NO: 58)	61-67-CPK063-G (SEQ ID NO: 59)	ERG19
2	100 ng each of $LEU2^{1096\ to\ 1770}$ and HISMX purified PCR products	61-67-CPK021-G (SEQ ID NO: 52)	61-67-CPK024-G (SEQ ID NO: 51)	$HISMX_LEU2^{1096\ to\ 1770}$
	100 ng each of ERG8 and P_{GAL} purified PCR products	61-67-CPK041-G (SEQ ID NO: 54)	61-67-CPK062-G (SEQ ID NO: 57)	$ERG8\text{-}P_{GAL}$
3	100 ng of $LEU2^{-100\ to\ 450}$ and HISMX- $LEU2^{1096\ to\ 1770}$ purified PCR products	61-67-CPK019-G (SEQ ID NO: 48)	61-67-CPK024-G (SEQ ID NO: 51)	**$LEU2^{-100\ to\ 450}$_HISMX_$LEU2^{1096\ to\ 1770}$**
	100 ng each of ERG8-P_{GAL} and ERG19 purified PCR products	61-67-CPK041-G (SEQ ID NO: 54)	61-67-CPK046-G (SEQ ID NO: 58)	**P_{GAL10}-ERG8_P_{GAL1}-ERG19**
** The HISMX marker in pAM330 originated from pFA6a-HISMX6-PGAL1 as described by van Dijken *et al.* ((2000) Enzyme Microb. Technol. 26(9-10):706-714).				

Vector pAM1419 was generated by removing from vector pTrc99A (Amman *et al.*, Gene 40:183-190 (1985)) two NcoI restriction sites. The NcoI restriction site in the multiple cloning site of vector pTrc99A was first changed to a NdeI restriction site using the Quikchange Multi Site-Directed Mutagenesis Kit (Stratagene Cloning Systems, La Jolla, CA) according to manufacturer recommended protocols. A second NdeI restriction site outside of the multiple cloning site at position 2699 of vector pTrc99A was then also removed using the Quikchange Multi Site-Directed Mutagenesis Kit, yielding vector pAM1419 (SEQ ID NO: 60).

Expression plasmid pAM1421 (FIG. 13A) were generated by inserting into vector pAM1419 the FS_S2D-Ec coding sequence. Vector pAM1419 was digested using BamHI and NdeI restriction endonucleases, the approximately 4.15 kb linearized vector backbone was gel purified, and the purified DNA fragment was digested with Calf Intestinal Alkaline Phosphatse (CIP) to remove 5' phosphate groups. The b-farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398; Picaud, et al, 2005) comprising an amino acid substitution at position 2 from serine to aspartate (S2D) and codon-optimized for expression in *Escherichia coli (*FS_S2D_Ec coding sequence; SEQ

ID NO: 61) with flanking BamHI and NdeI restriction sites was PCR amplified from other expression plasmids, the PCR product was digested using BamHI and NdeI restriction endonucleases, and then gel purified. Purified linearized vector and digested PCR product were ligated using T4 DNA ligase, yielding expression plasmid pAM1421.

Expression plasmid pAM353 was generated by inserting into the pRS425-Gal1 vector (Mumberg et. al. (1994) Nucl. Acids. Res. 22(25): 5767-5768) the coding sequence of the β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) codon-optimized for expression in *Saccharomyces cerevisiae* (FS_Aa_Sc coding sequence; SEQ ID NO: 68). The FS_Aa_Sc coding sequence was generated synthetically and was flanked by 5' BamHI and 3' XhoI restriction sites such that it could be cloned into compatible restriction sites of a cloning vector such as a standard pUC or pACYC origin vector. The FS_Aa_Sc coding sequence was excised again by digesting the construct using BamHI and XhoI restriction endonucleases, the approximately 1.7 kb DNA fragment comprising the FS_Aa_Sc coding sequence was gel purified, and the purified DNA fragment was ligated into the BamHI XhoI restriction site of the pRS425-Gal1 vector, yielding expression plasmid pAM353.

Expression plasmid pAM404 (FIG. 13B) was generated by inserting into vector pAM178 (SEQ ID NO: 69) the FS_Aa_Sc coding sequence. Vector pAM178 was digested using BamHI and NheI restriction endonucleases, and the approximately 7.3 kb vector backbone was gel purified. The FS_Aa_Sc coding sequence with flanking BamHI and NheI restriction sites was PCR amplified from pAM353, the PCR product was digested using BamHI and NheI restriction endonucleases, and the approximately 1.7 kb DNA fragment comprising the FS_Aa_Sc coding sequence was gel purified. The two gel purified DNA fragments were ligated, yielding expression plasmid pAM404.

The expression plasmids shown in Table 6 and FIG. 13B were generated by replacing in expression plasmid pAM404 the FS_Aa_Sc coding sequence with the indicated coding sequences. Vector pAM404 was digested using BamHI and NheI restriction endonucleases, and the approximately 7.3 kb linearized vector backbone lacking the FS_Aa_Sc coding sequence was gel purified. The FS_S2D_Ec coding sequence was PCR amplified from another expression plasmid using primers that overlap with terminal sequences of the linearized pAM404 vector backbone. The IS_Pn_Sc and TDS_Pn_Sc coding sequence were extracted by restriction endonuclease digestion from other expression vectors. The DNA fragments were gel purified. The FS_S2D_Ec coding

sequence was inserted into the purified linearized vector via homologous recombination by transforming strainY539 and selecting host cell transformants on Complete Synthetic Medium (CSM) lacking leucine (CSM-L) with 2% glucose as a sole carbon source. The IS_Pn_Sc and TDS_Fs_Sc coding sequences were ligated into the purified linearized vector using T4 DNA ligase.

Table 6 – Construction of pAM404-derived expression plasmids

Expression Plasmid	Coding Sequence	Size of Coding Sequence (bp)
pAM1765	FS_S2D_Ec (SEQ ID NO: 61)	1.7
pAM1795	TDS_Fs_Sc (SEQ ID NO: 70)	1.1
pAM1549	IS_Pn_Sc (SEQ ID NO: 71)	1.8

FS_S2D_Ec = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398; Picaud, et al, 2005) comprising an amino acid substitution at position 2 from serine to aspartate (S2D) and codon-optimized for expression in *Escherichia coli*

TDS_Fs_Sc = coding sequence of the trichodiene synthase gene of *Fusarium sporotrichioides* (GenBank accession number AF364179; *Holn, et al, 1989*) codon-optimized for expression in *Saccharomyces cerevisiae*

IS_Pn_Sc = isoprene synthase gene of *Populus nigra* (GenBank accession number AM410988; *Fortunati, et al, 2008*) codon-optimized for expression in *Saccharomyces cerevisiae*

Expression plasmid pAM1812 (SEQ ID NO: 72; FIG. 13C) was generated by inserting into expression plasmid pAM404 the TDS_Fs_Sc coding sequence. Expression plasmid pAM404 was digested using NotI restriction endonuclease, and the approximately 9.0 kb linearized plasmid was gel purified. The TDS_Fs_Sc coding sequence was PCR amplified from expression plasmid pAM1795 using primers that overlap with terminal sequences of the linearized pAM404, and the 1.2 kb PCR product comprising the TDS_Fs_Sc coding sequence was gel purified. The two gel purified DNA fragments were ligated via homologous recombination by transforming strain Y539 with both purified DNA fragments and selecting host cell transformants on CSM-L with 2% glucose as a sole carbon source.

The expression plasmids shown in Table 7 and FIG. 13C were generated by replacing in expression plasmid pAM1812 the FS_Aa_Sc coding sequence with the indicated coding sequences. Expression plasmid pAM1812 was digested using BamHI and NheI restriction endonucleases, and the approximately 7.2 kb linear plasmid lacking the FS_Aa_Sc coding sequence was gel purified. The coding sequences were extracted by BamHI and NheI restriction endonuclease digestion from other expression vectors, and the DNA fragment comprising the coding sequence were gel purified. The purified DNA fragments were finally ligated using T4 DNA ligase, yielding the expression plasmids.

Table 7 – Construction of pAM1812-derived expression plasmids		
Expression Plasmid	Coding Sequence	Size of Coding Sequence (kb)
pAM1895	FS_S2D_Ec (SEQ ID NO: 61)	1.7
pAM1896	FS_Ad_Sc (SEQ ID NO: 73)	2.3
pAM1948	IS_Pn_Sc (SEQ ID NO: 71)	1.8
FS_S2D_Ec = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) comprising an amino acid substitution at position 2 from serine to aspartate (S2D) and codon-optimized for expression in *Escherichia coli* FS_Ad_Sc = α–farnesene synthase of *Actinidia deliciosa* (GenBank Accession No. FJ265785; *Nieuwenhuizen, et al, 2009*) codon-optimized for expression in *Saccharomyces cerevisiae* IS_Pn_Sc = isoprene synthase of *Populus nigra* (Accession No. AM410988; *Fortunati, et al, 2008*) codon-optimized for expression in *Saccharomyces cerevisiae*		

Expression plasmid pAM1813 (FIG. 13D) was generated by inserting into expression plasmid pAM1795 the coding sequence of the FS_Aa_Sc coding sequence. Expression plasmid pAM1795 was digested using NotI restriction endonuclease, and the approximately 8.4 kb linearized plasmid was gel purified. The FS_Aa_Sc coding sequences was PCR amplified using primers that overlap with terminal sequences of the linearized pAM404, and the PCR product comprising the FS_Aa_Sc coding sequences was gel purified. The purified PCR product was ligated via homologous recombination by transforming it into strain Y539 and selecting host cell transformants on CSM-L with 2% glucose as a sole carbon source. Note that pAM1812 and pAM1813 are identical except that the promoters for TDS and FS are switched. Using promoters of different strengths allows variation of the farnesene/trichodiene ratios.

Expression plasmid pAM1653 was generated by inserting into vector pRS415 the IS_Pn_Sc coding sequence. Expression plasmid pAM1549 was digested using SapI restriction endonuclease, was treated with Klenow fragment to generated blunt ends, and digested again using NotI restriction endonuclease, and the 2.8 kb DNA fragment comprising the IS_Pn_Sc coding sequence and PGAL sequence was gel purified. Vector pRS415 was digested using NotI and AleI restriction endonucleases, and the 6.0 kb linearized vector backbone was gel purified. The two purified DNA fragments were ligated, yielding expression plasmid pAM1653.

Expression plasmid pAM1734 was generated by eliminating certain restriction sites from the multicloning region of expression plasmid pAM1653. Expression plasmid pAM1653 was digested using XbaI and HindIII restriction endonucleases, treated with Klenow fragment to generate blunt ends, and finally self-ligated, yielding expression plasmid pAM1734.

Expression plasmid pAM1764 (SEQ ID NO: 74) was generated by inserting into expression plasmid pAM1734 the FS_S2D_Ec coding sequence. Expression vector pAM1734 was digested using BamHI and NheI restriction endonucleases, and the approximately 6.9 kb linearized plasmid was gel purified. The FS_S2D_Ec coding sequence was PCR amplified from expression plasmid pAM1421 using primers that overlap with terminal sequences of the linearized pAM1734, and the 1.7 kb PCR product comprising the FS_S2D-Sc coding sequence was gel purified. The two purified DNA fragments were ligated via homologous recombination by transforming strain Y539 with both purified DNA fragments and selecting host cell transformants on CSM-L with 2% glucose as a sole carbon source.

Expression plasmid pAM1668 was generated by deleting from expression plasmid pAM1419 the lacI gene. Expression plasmid pAM1419 was digested using EcoRV and SapI restriction endonucleases, the digested plasmid was repaired using the End-It DNA End-Repair Kit (Epicentre, Madison, WI) according to manufacturer recommended protocols, and the end-repaired vector was self-ligated, yielding vector pAM1668.

The expression plasmids shown in Table 8 and FIG. 13E were generated by inserting into vector pAM1668 the indicated coding sequences. Vector pAM1668 was digested using BamHI and NdeI restriction endonucleases, and the approximately 2.9 kb linearized vector backbone was gel purified. The coding sequences were PCR amplified using primers that overlap with terminal sequences of the linearized pAM1668, the PCR products were digested using BamHI and NdeI restriction endonucleases, and the digested PCR products comprising the coding sequences were gel purified. Purified linearized vector and digested PCR products were finally ligated, yielding the expression plasmids.

Table 8 – Construction of pAM1688-derived expression plasmids		
Expression Plasmid	Coding Sequence	Size of Coding Sequence (bp)
pAM1670	FS_S2D_Ec (SEQ ID NO: 61)	1725
pAM2158	FS_S2D_Sc (SEQ ID NO: 76)	1725
pAM2157	FS_ Aa_Ec (SEQ ID NO: 62)	1725
pAM2117	FS_Aa_Sc (SEQ ID NO: 68)	1740
pAM2096	FS_Cj_Ec (SEQ ID NO: 63)	1686
pAM2097	FS_Cs_Ec (SEQ ID NO: 66)	1686
pAM2098	FS_Pt_Ec (SEQ ID NO: 67)	1725
pAM2101	PS_S_Ec (SEQ ID NO: 64)	1014
pAM2104	TDS_Fs_Ec (SEQ ID NO: 65)	1125

FS_S2D_Ec = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) comprising an amino acid substitution at position 2 from serine to aspartate (S2D) and codon-optimized for expression in *Escherichia coli*
FS_S2D_Sc = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) comprising an amino acid substitution at position 2 from serine to aspartate (S2D) and codon-optimized for expression in *Saccharomyces cerevisiae*
FS_Aa_Sc = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) codon-optimized for expression in *Saccharomyces cerevisiae*
FS_Aa_Ec = β–farnesene synthase gene of *Artemisia annua* (GenBank accession number AY835398) codon-optimized for expression in *Escherichia coli*
FS_Cj_Ec = β–farnesene synthase of *Citrus junos* (GenBank Accession No. AAK54279; *Maruyama, et al, 2001*) codon-optimized for expression in *Escherichia coli*
FS_Cs_Ec = α–farnesene synthase of *Cucumis sativus* (GenBank Accession No. AAU05951; *Mercke, et al, 2004*) codon-optimized for expression in *Escherichia coli*
FS_Pt_Ec = α–farnesene synthase of *Pinus taeda* (GenBank Accession No. AAO61226; Phillips, et al, 2003) codon-optimized for expression in *Escherichia coli*
PS_S_Ec = pentalenene synthase of *Streptomyces sp.* (GenBank Accession No. AAA19131; Cane, et al, 1994) codon-optimized for expression in *Escherichia coli*
TDS_Fs_Ec = trichodiene synthase of *Fusarium sporotrichioides* (GenBank accession number AF364179; *Holn, et al, 1989*) codon-optimized for expression in *Escherichia coli*

Construct A was generated by PCR amplification as described in Table 9. The construct comprises the upstream region of the NDT80 gene of *Saccharomyces cerevisiae* (NDT80 nucleotide positions -175 to -952), the LEU2 marker of Saccharomyces cerevisiae (LEU2 nucleotide positions -661 to +1541), the promoter of the GAL1 gene of Saccharomyces cerevisiae (GAL1 nucleotide positions -1 to -667), the FS_Aa_Sc coding sequence, the terminator of the CYC1 gene of *Saccharomyces cerevisiae* (CYC1 nucleotide positions +331 to +521), and the downstream region of the NDT80 gene of *Saccharomyces cerevisiae* (NDT80 nucleotide positions +1685 to +2471). FIG. 1W shows a map and SEQ ID NO: 86 the nucleotide sequence of Construct A.

Table 9 – PCR reactions performed to generate Construct A				
PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y002 genomic DNA	PW-091-144-CPK640 (SEQ ID NO: 77)	AM-288-90-CPK1615 (SEQ ID NO: 78)	$NDT80^{-175\ to\ -952}$
		AM-288-90-CPK1620 (SEQ ID NO: 79)	PW-091-144-CPK649 (SEQ ID NO: 80)	$NDT80^{+1685\ to\ +2471}$
		AM-288-90-CPK1616 (SEQ ID NO: 81)	AM-288-90-CPK1638 (SEQ ID NO: 82)	$GAL1^{-1\ to\ -667}$
	10 ng of plasmid pAM404	AM-288-90-CPK1618 (SEQ ID NO: 83)	AM-288-90-CPK1619 (SEQ ID NO: 84)	FS_Aa_Sc
		AM-288-90-CPK1639 (SEQ ID NO: 85)	AM-288-90-CPK1621 (SEQ ID NO: 75)	$CYC1^{+331\ to\ +521}$
	10 ng of plasmid pRS415 (*Sikorski, et al, 1989*)	AM-288-90-CPK1614 (SEQ ID NO: 87)	AM-288-90-CPK1617 (SEQ ID NO: 88)	$LEU2^{-661\ to\ +1541}$
2	25 ng of $CYC1^{+331\ to\ +521}$ and equimolar amounts of each of $NDT80^{-175\ to\ -952}$, $LEU2^{-661\ to\ +1541}$, $GAL1^{-1\ to\ -667}$, FS_Aa_Sc, and $NDT80^{+1685\ to\ +2471}$ gel purified PCR products	PW-091-144-CPK640 (SEQ ID NO: 77)	PW-091-144-CPK649 (SEQ ID NO: 80)	Construct A
PCR amplifications were done using the Phusion High Fidelity DNA Polymerase System (Finnzyme, Inc., Espoo, Finland). PCR products were gel purified using the E.Z.N.A. ® Gel Extraction Kit (Omega Bio-Tek Inc., Norcross, GA) according to manufacturer's suggested protocols.				

Construct B (FIG. 1X) was generated by PCR amplification. The construct comprises the upstream region of the NDT80 gene of *Saccharomyces cerevisiae* (NDT80 nucleotide positions -175 to -952), the LEU2 marker of *Saccharomyces cerevisiae* (LEU2 nucleotide positions -661 to +1541), the promoter of the GAL1 gene of *Saccharomyces cerevisiae* (GAL1 nucleotide positions -1 to -667), the HISG marker, the terminator of the CYC1 gene of *Saccharomyces cerevisiae* (CYC1 nucleotide positions +331 to +521), and the downstream region of the NDT80 gene of *Saccharomyces cerevisiae* (NDT80 nucleotide positions +1685 to +2471).

Construct D was generated by PCR amplification as described in Table 10. The construct comprises the FS_A_5.3 coding sequence flanked by the promoter of the GAL1 gene of *Saccharomyces cerevisiae* (PGAL1; GAL1 nucleotide positions -1 to -455) and the terminator of the PGK1 gene of *Saccharomyces cerevisiae* (TPGK1; PGK1 nucleotide positions +1159 to +1547), and the TDS_Fs_Sc coding sequence flanked by the promoter of the GAL10 gene of *Saccharomyces cerevisiae* (PGAL10; GAL10 nucleotide positions -1 to -202) and the terminator of the ADH1 gene of *Saccharomyces cerevisiae* (TADH1;

ADH1 nucleotide positions -1 to -166). FIG. 1Z shows a map and SEQ ID NO: 116 the nucleotide sequence of Construct D.

Table 10 – PCR reactions performed to generate Construct D

PCR Round	Template	Primer 1	Primer 2	PCR Product
1	100 ng of Y002 genomic DNA	AM-288-160-CPK2039 (SEQ ID NO: 123)	AM-288-160-CPK2041 (SEQ ID NO: 125)	TPGK1
	10 ng of a plasmid comprising the FS_A_5.3 coding sequences	AM-288-160-CPK2040 (SEQ ID NO: 124)	AM-288-110-CPK1903 (SEQ ID NO: 122)	FS_A_5.3
	10 ng of plasmid pAM1948	AM-288-160-CPK2045 (SEQ ID NO: 129)	AM-288-160-CPK2046 (SEQ ID NO: 130)	PGAL10_TDS_TADH1
	10 ng of plasmid comprising spacer	AM-288-160-CPK2042 (SEQ ID NO: 126)	AM-288-160-CPK2044 (SEQ ID NO: 128)	spacer
	10 ng of plasmid pAM404	AM-288-110-CPK1849 (SEQ ID NO: 121)	AM-288-160-CPK2043 (SEQ ID NO: 127)	PGAL1
2	25 ng of spacer and equimolar amounts of each of PGAL1, FSA_A_5.3, TPGK1, and PGAL10_TDS_TADH1 gel purified PCR products	AM-288-160-CPK2039 (SEQ ID NO: 123)	AM-288-160-CPK2046 (SEQ ID NO: 130)	Construct D

PCR amplifications were done using the Phusion High Fidelity DNA Polymerase System (Finnzyme, Inc., Espoo, Finland). PCR products were gel purified using the E.Z.N.A.®Gel Extraction Kit (Omega Bio-Tek Inc., Norcross, GA) according to manufacturer's suggested protocols.

Expression plasmid pAM2191 was created by in transforming exponentially growing Y 3198 cells with 100 ng of vector pAM552 (SEQ ID NO: 156) digested using FastDigest® BstZ17I restriction enzyme (Fisher Scientific Worldwide, Hampton, NH) and 300 ng of Construct D. Host cell transformants were plated on CSM-L agar plates with 2% glucose as a sole carbon source, and the plates were incubated for 3 days at 30°C until individual colonies were ~1mm in diameter. DNA was harvested from these colonies using the Zymoprep™ Yeast Plasmid Miniprep Kit II (Zymo Research Corporation, Orange, CA), and the harvested DNA was transformed into the chemically competent XL1Blue Escherichia coli (Agilent Technologies Inc. , Santa Clara, CA). Host cell transformants were plated to Lysogeny broth agar media supplemented with carbenicillin, and incubated for 24 hours at 37°C until individual colonies were visible. Plasmid DNA was harvested from these colonies using the QIAprep Spin Miniprep Kit (QIAGEN Inc,

Valencia, CA), and the plasmid DNA was sequenced to confirm correct creation of expression plasmid pAM2191.

Example 2

This example describes methods for making yeast strains useful in the generation and characterization of terpene synthases variants.

Strains Y93 (MAT A) and Y94 (MAT alpha) were generated by replacing the promoter of the ERG9 gene of yeast strains Y002 and Y003 (CEN.PK2 background MAT A or MAT alpha, respectively; ura3-52; trp1-289; leu2-3,112; his3Δ1; MAL2-8C; SUC2; van Dijken *et al.* (2000) Enzyme Microb. Technol. 26:706-714), respectively, with the promoter of the MET3 gene of *Saccharomyces cerevisia.* To this end, exponentially growing Y002 and Y003 cells were transformed with integration construct i8 (SEQ ID NO: 87), which comprised the kanamycin resistance marker (KanMX) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis,* the ERG9 coding sequence, a truncated segment of the ERG9 promoter (trune. PERG9), and the MET3 promoter (PMET3), flanked by ERG9 upstream and downstream sequences (FIG. 1A). Host cell transformants were selected on medium comprising 0.5 ug/mL Geneticin (Invitrogen Corp., Carlsbad, CA), and selected clones were confirmed by diagnostic PCR, yielding strains Y93 and Y94.

Strains Y176 (MAT A) and Y177 (MAT alpha) were generated by replacing the coding sequence of the ADE1 gene in strains Y93 and Y94, respectively, with the coding sequence of the LEU2 gene of *Candida glabrata* (*Cg*LEU2). To this end, the 3.5 kb *Cg*LEU2 genomic locus was PCR amplified from *Candida glabrata* genomic DNA (ATCC, Manassas, VA) using primers 61-67-CPK066-G (SEQ ID NO: 88) and 61-67-CPK067-G (SEQ ID NO: 89), and transforming the PCR product into exponentially growing Y93 and Y94 cells. Host cell transformants were selected on CSM-L, and selected clones were confirmed by diagnostic PCR, yielding strains Y176 and Y177.

Strain Y188 was generated by introducing into strain Y176 an additional copy of the coding sequences of the ERG13, ERG10, and ERG12 genes of *Saccharomyces cerevisia*, and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisia*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisia*. To this end, exponentially growing Y176 cells were transformed with 2 μg of expression plasmids pAM491 and pAM495 digested with PmeI restriction endonuclease (New England Biolabs, Beverly, MA). Host cell

transformants were selected on CSM lacking uracil and histidine (CSM-U-H), and selected clones were confirmed by diagnostic PCR, yielding strain Y188.

Strain Y189 was generated by introducing into strain Y177 an additional copy of the coding sequences of the ERG20, ERG8, and ERG19 genes of *Saccharomyces cerevisia*, and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisia*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisia*. To this end, exponentially growing Y188 cells were transformed with 2 ug of expression plasmids pAM489 and pAM497 digested with PmeI restriction endonuclease. Host cell transformants were selected on CSM lacking tryptophan and histidine (CSM-T-H), and selected clones were confirmed by diagnostic PCR, yielding strain Y189.

Strain Y238 was generated by mating strains Y188 and Y189, and by introducing an additional copy of the coding sequence of the IDI1 gene of *Saccharomyces cerevisia* and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisia*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisia*. To this end, approximately 1 x 10^7 cells of strains Y188 and Y189 were mixed on a YPD medium plate for 6 hours at room temperature, diploid cells were selected on CSM-H-U-T, and exponentially growing diploids were transformed with 2 ug of expression plasmid pAM493 digested with PmeI restriction endonuclease. Host cell transformants were selected on CSM lacking adenine (CSM-A), and selected clones were confirmed by diagnostic PCR, yielding strain Y238.

Strains Y210 (MAT A) and Y211 (MAT alpha) were generated by sporulating strain Y238. The diploid cells were sporulated in 2% potassium acetate and 0.02% raffinose liquid medium, and approximlately 200 genetic tetrads were isolated using a Singer Instruments MSM300 series micromanipulator (Singer Instrument Co, LTD. Somerset, UK). Spores were selected on CSM-A-H-U-T, and selected clones were confirmed by diagnostic PCR, yielding strains Y210 (MAT A) and Y211 (MAT alpha).

Strain Y221 was generated by transforming exponentially growing Y211 cells with vector pAM178. Host cell transformants were selected on CSM-L.

Strain Y290 was generated by deleting the coding sequence of the GAL80 gene of strain Y221. To this end, exponentially growing Y221 cells were transformed with integration construct i32 (SEQ ID NO: 90), which comprised the hygromycin B resistance marker (hph) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis* flanked by GAL80 upstream and downstream sequences (FIG. 1B). Host cell

transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y290.

Strain Y318 was generated by removing the pAM178 vector from strain Y290 by serial propagation in leucine-rich media, and testing individual colonies for their inability to grow on CSM-L, yielding strain Y318.

Strain Y409 was generated by introducing a heterologous nucleotide sequence encoding a β-farnesene synthase into strain Y318. To this end, exponentially growing Y318 cells were transformed with expression plasmid pAM404. Host cell transformants were selected on CSM-L, yielding strain Y409.

Strain Y419 was generated by rendering the GAL promoters of strain Y409 constitutively active. To this end, exponentially growing Y409 cells were transformed with integration construct i33 (SEQ ID NO: 91), which comprised the nourseothricin resistance marker of *Streptomyces noursei* (NatR) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, and the coding sequence of the GAL4 gene of *Saccharomyces cerevisiae* under regulatory control of an "operative constitutive" version of its native promoter (PGAL4oc; Griggs & Johnston (1991) *PNAS* **88**(19):8597-8601) and the GAL4 terminator (TGAL4) (FIG. 1C), flanked by upstream and downstream sequences of the modified ERG9 promoter and coding sequences. Host cell transformants were selected on medium comprising nourseothricin, and selected clones were confirmed by diagnostic PCR, yielding strain Y419.

Strain Y677 was generated by introducing at the modified GAL80 locus of strain Y419 an additional copy of the coding region of the ERG12 gene of *Saccharomyces cerevisiae* under regulatory control of the promoter of the GAL1 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y677 cells were transformed with integration construct i37 (SEQ ID NO: 92), which comprised the kanamycin resistance marker of *Streptomyces noursei* (KanR) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, and the coding and terminator sequences of the ERG12 gene of *Saccharomyces cerevisiae* flanked by the GAL1 promoter (PGAL1) and the ERG12 terminator (TERG12) (FIG. 1D). Host cell transformants were selected on medium comprising kanamycin, and selected clones were confirmed by diagnostic PCR, yielding strain Y677.

Strain Y1551 was generated from strain Y677 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y1551.

Strain Y1778 was generated from strain Y1551 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y1778.

Strain Y1816 was generated by replacing the HXT3 coding sequence of strain Y1778 with two copies of an acetoacetyl-CoA thiolase coding sequence, one being derived from *Saccharomyces cerevisiae* and the other from *C. butylicum*, and one copy of the coding sequence of the HMGS gene of *B. juncea*. To this end, exponentially growing Y1778 cells were transformed with integration construct i301 (SEQ ID NO: 93), which comprised the hygromycin B resistance marker (hyg) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, the coding sequence of the ERG10 gene of *Saccharomyces cerevisiae* flanked by a truncated TDH3 promoter (tPTDH3) and the AHP1 terminator (TAHP1), the coding sequence of the acetoacetyl-CoA thiolase gene of *C. butylicum* (thiolase) flanked by the YPD1 promoter (PYPD1) and CCW12 terminator (TCCW12), and the coding sequence of the HMGS gene of *B. juncea* (HMGS) preceded by the TUB2 promoter (PTUB2), flanked by upstream and downstream sequences of the HXT3 gene of *Saccharomyces cerevisiae* (FIG. 1E). Host cell transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y1816.

Strain Y2055 was generated from strain Y1778 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2055.

Strain Y2295 was generated from strain Y2055 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2295.

Strain Y3111 was generated by switching the mating type of strain Y2295 from MAT A to MAT alpha. To this end, exponentially growing Y2295 cells were transformed with integration construct i476 (SEQ ID NO: 94), which comprised the MAT alpha mating locus and the hygromycin B resistance marker (hygA) (FIG. 1F). Host cell transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y3111.

Strain Y2168 was generated from strain Y1816 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2168.

Strain Y2446 was generated from strain Y2168 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2446.

Strain Y3118 was generated by inserting into the native URA3 locus of strain Y2446 the coding sequence, promoter, and terminator of the GAL80 gene of

Saccharomyces cerevisiae. To this end, exponentially growing Y2446 cells were transformed with integration construct i477 (SEQ ID NO: 95), which comprised the promoter, terminator, and coding sequence of the GAL80 gene of *Saccharomyces cerevisiae* (GAL80) flanked by overlapping URA3 sequences (which enable loop-out excision of the GAL80 gene by homologous recombination and restoration of the original URA3 sequence) (FIG. 1G). Host cell transformants were selected on medium comprising 5-FOA, yielding strain Y3118.

Strain Y3125 was generated from strain Y3118 by removing expression plasmid pAM404. To this end, strain Y3118 was first incubated in 3 mL YPD+0.5% leucine medium with 40 mg/L uracil (YPD+L+U). Cells were diluted by 100-fold in fresh YPD+L+U every 24 hours for up to 5 days to lose expression plasmid pAM404, and were then plated on YPD and incubated at 30°C for up to 5 days. Small colonies were replica-plated on both YPD and CSM-L, and colonies that could grow on YPD but failed to grow on CSM-L were identified, yielding strain Y3125.

The strains shown in Table 11 were generated by transforming exponentially growing Y3125 cells with the indicated expression plasmids comprising the indicated coding sequences, and selecting host cell transformants on CSM-L.

Table 11 – FS and TDS expressing yeast strains

Strain	Plasmid	Coding sequence (promoter)
Y3353	pAM1812	FS_Aa_Sc (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})
Y3354	pAM1813	FS_Aa_Sc (P_{GAL10}) TDS_Fs_Sc (P_{GAL1})
Y3394	pAM1895	FS_S2D_Ec (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})
Y3395	pAM1896	FS_Ad_Sc (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})

Strain Y227 was generated by transforming strain Y211 with expression plasmid pAM426. Host cell transformants were selected on CSM-L.

Strain Y3198 was generated by removing expression plasmid pAM426 from strain Y227. To this end, Y227 cells were propagated in YPD+L for 4 days. Every 24 hours, the culture was inoculated to an OD600 of 0.05 in fresh YPD+L. After 4 days, cells were serial diluted and plated onto YPD solid agar, and the plates were incubated at 30°C for 4 days. Smaller colonies were replica-plated on both YPD and CSM-L, and colonies that could grow on YPD but failed to grow on CSM-L were identified, yielding strain Y3198.

Strain Y3215 was generated by mating strains Y3111 and Y3118. Approximately 1 x 10^7 cells of strains Y3111 and Y3118 were mixed on a YPD medium plate for 6 hours at room temperature to allow for mating, followed by plating on YPD agar plate to isolate single colonies. Diploids were identified by screening by colony PCR for the presence of both the hphA-marked MAT alpha locus and the wild-type MAT A locus.

Strain Y3000 was generated by sporulating strain Y3215 and looping out the GAL80 coding sequence. The diploid cells were sporulated in 2% potassium acetate and 0.02% raffinose liquid medium. Random spores were isolated, plated on YPD agar, grown for 3 days, and then replica-plated to CSM-U to permit growth only of cells lacking GAL80 (*i.e.*, having a functional URA3 gene). Spores were then tested for β–farnesene production, the best producer was identified, and the presence of integration construct i301 was confirmed by diagnostic PCR, yielding strain Y3000.

Strain Y3284 was generated by removing the URA3 marker from strain Y3000. To this end, exponentially growing Y3000 cells were transformed with integration construct i94 (SEQ ID NO: 96), which comprised the hisG coding sequence of *Salmonella,* and the coding sequence of the ERG13 gene and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae* under control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*, flanked by upstream and downstream sequences of the URA3 gene of *Saccharomyces cerevisiae* (FIG. 1H). Host cell transformants were selected on medium comprising 5-FOA, and selected clones were confirmed by diagnostic PCR, yielding strain Y3284.

Strain Y3385 was generated by replacing the NDT80 coding sequence of strain Y3284 with an additional copy of the coding sequence of an acetyl-CoA synthetase gene of *Saccharomyces cerevisiae* and the coding sequence of the PDC gene of *Z. mobilis*. To this end, exponentially growing Y3385 cells were transformed with integration construct i467 (SEQ ID NO: 97), which comprised the URA3 marker, the coding sequence of the ACS2 gene of *Saccharomyces cerevisiae* (ACS2) flanked by the HXT3 promoter (PHXT3) and PGK1 terminator (TPGK1), and the coding sequence of the PDC gene of *Z. mobilis* (zmPDC) flanked by the GAL7 promoter (PGAL7) and the TDH3 terminator (TTDH3), flanked by upstream and downstream NDT80 sequences (FIG. 1I). Host cell transformants were selected on CSM-U, and selected clones were confirmed by diagnostic PCR, yielding strain Y3385.

Strain Y3547 was generated from strain Y3385 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y3547.

Strain Y3639 was generated from strain Y3547 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y3639.

Strain Y3902 was generated by disrupting the URA3 marker at the NDT80 locus of strain Y3639 with an integration of the coding sequence, promoter, and terminator of the GAL80 gene of *Saccharomyces cerevisiae*. In this instance, a mutant allele of GAL80 called $GAL80^{s-2}$ was used (Nucleic Acids Research (1984) 12(24):9287-9298). Exponentially growing Y3639 cells were transformed with integration construct i601 (SEQ ID NO: 98), which comprised the promoter, terminator, and coding sequence of the GAL80 gene of *Saccharomyces cerevisiae* (PGAL80, TGAL80, and GAL80, respectively) flanked by overlapping URA3 sequences (which enable loop-out excision of the GAL80 gene by homologous recombination and restoration of the original URA3 sequence), the coding sequences of the ACS2 gene of *Saccharomyces cerevisiae* (ACS2) flanked by the HXT3 promoter (PHXT3) and the PGK1 terminator (TPGK1), and the PDC coding sequence of *Z. mobilis* (zmPDC) flanked by the GAL7 promoter (PGAL7) and the TDH3 terminator (TTDH3), flanked by upstream and downstream NDT80 sequences (FIG. 1J). Host cell transformants were selected on medium comprising 5-FOA, and selected clones were confirmed by diagnostic PCR, yielding strain Y3902.

Strain Y4027 was generated from strain Y3902 by removing expression plasmid pAM404 by serial propagation in leucine-rich media, and testing individual colonies for their inability to grow on medium lacking leucine.

Strain Y4909 was generated by replacing the ADH5 coding sequence of strain Y4027 with the FS_D_3.5 coding sequence and the FS_C_7 coding sequence (see Table 17). To this end, exponentially growing Y4909 cells were transformed with integration construct i2125 (SEQ ID NO: 99), which comprised the LEU2 coding sequence (LEU2) and the farnesene synthase variant coding sequences flanked by GAL1 or GAL10 promoter (PGAL1 or PGAL10) and the CYC1 or ADH1 terminator (TCYC1 or TADH1), respectively, flanked by upstream and downstream ADH5 sequences (FIG. 1K). Host cell transformants were selected on CSM-L, and selected clones were confirmed by diagnostic PCR, yielding strain Y4909.

Strain Y4959 was generated by removing from strain Y4909 the coding sequence, promoter, and terminator of the GAL80 gene of *Saccharomyces cerevisiae*. To this end,

Y4909 cells were plated on CSM-U to select for spontaneous GAL80 "loop-out" recombination events, yielding strain Y4959.

Strain Y5444 was generated by inserting at the CAN1 locus of strain Y4959 the TDS_Fs_Sc coding sequence. To this end, exponentially growing Y4959 cells were transformed with integration construct i2608 (SEQ ID NO: 100), which comprised the TDS_Fs_Sc coding sequence flanked by the GAL1 promoter of *Saccharomyces cerevisiae* (PGAL1) and the CYC1 terminator of *Saccharomyces cerevisiae* (TCYC1) (FIG. 1L). Host cell transformants were selected on YNB medium comprising canavanine and lacking arginine, and selected clones were confirmed by diagnostic PCR, yielding strain Y5444.

Strain Y4910 was generated by inserting at the ADH5 locus of Y4027 the FS_A_5.3 coding sequence, the FS_B_5.3 coding sequence, the FS_C_7 coding sequence, and the FS_D_3.5 coding sequence (see Table 17). To this end, exponentially growing Y4909 cells were transformed with integration construct i2127 (SEQ ID NO: 101), which comprised the LEU2 coding sequence (LEU2) and the farnesene synthase variant coding sequences flanked by the GAL1 or GAL10 promoter (PGAL1 or PGAL10) and the ADH1 or CYC1 terminator (TADH1 or TCYC1), flanked by upstream and downstream ADH5 sequences (FIG. 1M), and which was generated by co-transformation with two overlapping segments of i2127, thereby stimulation homologous recombination between the two overlapping plasmid inserts. Host cell transformants were selected on CSM-L, and selected clones were confirmed by diagnostic PCR, yielding strain Y4910.

Strain Y4960 was generated by removing from strain Y4910 the coding sequence, promoter, and terminator of the GAL80 gene of *Saccharomyces cerevisiae*. To this end, Y4910 cells were plated on CSM-U to select for spontaneous GAL80 "loop-out" recombination events, yielding strain Y4960.

Strain Y5445 was generated by inserting at the CAN1 locus of strain Y4960 the TDS_Fs_Sc coding sequence. To this end, exponentially growing Y4959 cells were transformed with integration construct i2608 (SEQ ID NO: 100), which comprised the TDS_Fs_Sc coding sequence flanked by the GAL1 promoter of *Saccharomyces cerevisiae* (PGAL1) and the CYC1 terminator of *Saccharomyces cerevisiae* (TCYC1) (FIG. 1L). Host cell transformants were selected on YNB medium comprising canavanine and lacking arginine, and selected clones were confirmed by diagnostic PCR, yielding strain Y5445.

Strain Y5064 was generated by removing the URA3 marker from strain Y4960. To this end, exponentially growing Y4960 cells ware transformed with integration construct i569 (SEQ ID NO: 102), which comprised the coding sequence of the ACS2 gene of *Saccharomyces cerevisiae* (ACS2) flanked by the HXT3 promoter (PHXT3) and the PGK1 terminator (TPGK1), and the coding sequence of the PDC gene of *Z. mobilis* (zmPDC) flanked by the GAL7 promoter (PGAL7) and the TDH3 terminator (TTDH3), flanked by upstream and downstream NDT80 sequences (FIG. 1N). Host cell transformants were selected based on their ability to grow on medium containing 5-FO, and selected clones were confirmed by diagnostic PCR, yielding strain Y5064.

Strain Y5065 was generated by inserting at the BIO4 locus of strain Y5064 the FS_A_5.3 coding sequence and the FS_B_5.3 coding sequence (see Table 17). To this end, exponentially growing Y5064 cells were transformed with integration construct i2124 (SEQ ID NO: 103), which comprised the URA3 coding sequence (URA3), and the farnesene synthase variant coding sequences flanked by the GAL1 or GAL10 promoter (PGAL1 or PGAL10) and ADH1 or CYC1 terminator (TADH1 or TCYC1), flanked by upstream and downstream BIO4 sequences (FIG. 1O). Host cell transformants were selected CSM-U, and selected clones were confirmed by diagnostic PCR, yielding strain Y5065.

Strain Y5066 was generated by inserting at the BIO4 locus of strain Y5064 the FS_A_5.3 coding sequence, the FS_B_5.3 coding sequence, the FS_C_7 coding sequence, and the FS_D_3.5 coding sequence (see Table 17). To this end, exponentially growing Y5064 cells were transformed with integration construct i2127 (SEQ ID NO: 101), which comprised the LEU2 coding sequence (LEU2) and the farnesene synthase variant coding sequences flanked by the GAL1 or GAL10 promoter (PGAL1 or PGAL10) and the ADH1 or CYC1 terminator (TADH1 or TCYC1), flanked by upstream and downstream ADH5 sequences (FIG. 1M), and which was generated by co-transformation with two overlapping segments of i2127, thereby stimulation homologous recombination between the two overlapping plasmid inserts. Host cell transformants were selected on CSM-L, and selected clones were confirmed by diagnostic PCR, yielding strain Y5066.

Strain Y5446 was generated by inserting at the CAN1 locus of strain Y5065 the TDS_Fs_Sc coding sequence. To this end, exponentially growing Y4959 cells were transformed with integration construct i2608 (SEQ ID NO: 100), which comprised the TDS_Fs_Sc coding sequence flanked by the promoter of the GAL1 gene of *Saccharomyces cerevisiae* (PGAL1) and the terminator of the CYC1 gene of

Saccharomyces cerevisiae (TCYC1) (FIG. 1L). Host cell transformants were selected on YNB medium comprising canavanine and lacking arginine, and selected clones were confirmed by diagnostic PCR, yielding strain Y5446.

Strain Y5447 was generated by inserting at the CAN1 locus of strain Y5066 the TDS_Fs_Sc coding sequence. To this end, exponentially growing Y4959 cells were transformed with integration construct i2608 (SEQ ID NO: 100), which comprised the TDS_Fs_Sc coding sequence flanked by the GAL1 promoter of *Saccharomyces cerevisiae* (PGAL1) and the CYC1 terminator of *Saccharomyces cerevisiae* (TCYC1) (FIG. 1L). Host cell transformants were selected on YNB medium comprising canavanine and lacking arginine, and selected clones were confirmed by diagnostic PCR, yielding strain Y5447.

Strain Y224 was generated by introducing into strain Y211 an expression plasmid encoding an amorphadeine synthase (ADS), a cytochrome P450 monooxygenase (AMO, CYP71AV1), and a NADP-cytochrome P450 oxidoreductase (CPR). To this end, exponentially growing Y211 cells were transformed with expression plasmid pAM322, which comprised the coding sequence of the ADS gene of *Artemisia annua* (ADS), the coding sequence of the AMO gene of *Artemisia annua*, and the coding sequence of the CPR gene of *Artemisia annua,* all codon-optimized for expression in *Saccharomyces cerevisiae* and under regulatory control of the promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*. Host cell transformants were selected CSM-L, yielding strain Y224.

Strain Y284 was generated by replacing the divergent promoter of the GAL1 and GAL10 genes and the GAL1 coding sequence of strain Y224 with the hygromycin B resistance marker (hphA). To this end, exponentially growing Y224 cells were transformed with integration construct i65 (SEQ ID NO: 104; FIG. 1P). Host cell transformants were selected based on their resistance to hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y284.

Strain Y301 was generated by putting the ERG9 gene of strain Y284 under regulatory control of the promoter of the CTR3 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y284 cells were transformed with integration construct i10 (SEQ ID NO: 105), which comprised the coding sequence of a D-serine deaminase (dsdA) and the CTR3 promoter (PCTR3), flanked by upstream and coding sequences of the ERG9 gene (FIG. 1Q). Host cell transformants were selected based on their ability to grow on D-

serine deaminase, and selected clones were confirmed by diagnostic PCR, yielding strain Y301.

Strain Y539 was generated from strain Y301 by removing expression plasmid pAM322. To this end, strain Y301 was propagated in rich Yeast Peptone Dextrose (YPD) medium containing 0.5% leucine (w/v) for 4 days. Every 24 hours, the culture was inoculated to an OD600 of 0.05 in fresh YPD containing 0.5% leucine (w/v). After 4 days, cells were serial diluted and plated onto YPD solid agar, and the plates were incubated at 30˚C for 4 days. Two distinct colony sizes were observed. Smaller colonies (indicating a loss of pAM322) were replica plated onto minimal medium lacking leucine. A clone that was able to grow on YPD solid agar but was unable to grow on medium lacking leucine was selected as strain Y539.

Example 3

This example demonstrates the feasibility of using FPP starvation based selection in *Escherichia coli* to screen for terpene synthases with improved *in vivo* performance.

DH5α chemical- or electro-competent *Escherichia coli* cells (Invitrogen, Carlsbad, CA) were transformed with 5 ng of an expression plasmid selected from the group consisting of expression plasmids pAM1668 (negative control), pAM1670, pAM2096, pAM2097, pAM2098, pAM2101, and pAM2104. Host cell transformants were plated on agar plates comprising carbenicillin, and the plates were incubated at 30° C for 2-3 days.

As shown in FIG. 2, cells transformed with expression plasmid pAM2097 or pAM2096 produced colonies of similar size as cells transformed with empty vector (pAM1668). However, cells transformed with expression plasmid pAM1670, pAM2098, as well as cells transformed with expression plasmids pAM2104 or pAM2101 (data not shown), produced colonies that were smaller than those produced by the control. The smaller colony sizes were likely due to FPP starvation triggered by the conversion in the host cells of FPP into farnesene mediated by the expression of an active sesquiterpene synthase. The *in vivo* activities of these sesquiterpene synthases were confirmed by GC analysis of host cells at 72 hours in shake flasks (FIG. 4), showing that FPP starvation based selection in *Escherichia coli* can be used to screen sesquiterpene synthases for *in vivo* enzyme activity.

Example 4

This example demonstrates the feasibility of using FPP toxicity based growth selection in *Escherichia coli* to screen for sesquiterpene synthases with improved *in vivo* performance.

DH5α chemical- or electro-competent *Escherichia coli* cells (Invitrogen, Carlsbad, CA) were co-transformed with 5 ng of expression plasmid pAM765 (encodes enzymes of the MEV pathway that collectively increase production of FPP in the host cells) and 5 ng of an expression plasmid selected from the group consisting of expression plasmids pAM1668 (negative control), pAM1670, pAM2117, pAM2157, pAM2158, pAM2098, pAM2104, pAM2097, pAM2101, and pAM2096. Host cell transformants were plated on agar plates comprising chloramphenicol and carbenicillin, and the plates were incubated at 30°C for 2-5 days.

As shown in FIG. 3 and Table 12, cells transformed with empty vector (pAM1668) produced no colonies, likely due to cell death triggered by the accumulation of toxic FPP in the host cells (Withers at al. (2007) *Appl. Environ. Microbiol.* 73:6277-6283). Similarly, cells transformed with expression plasmid pAM2158 or pAM2117 did not form colonies, likely due to the fact that the farnesene synthase coding sequences of these plasmids were not codon-optimized for efficient expression in the *Escherichia coli* host cells. Cells transformed with expression plasmids pAM2096 also did not form colonies, suggesting that the *Citrus junos* farnesene synthase does not possess sufficient activity in *Escherichia coli* host cells. All other transformed cells produced colonies within 1-3 days of culture, presumably due to the conversion of FPP into less toxic sesquiterpenes by the active sesquiterpene synthases encoded by expression plasmids pAM1670, pAM2157, pAM2098, pAM2104, pAM2097, and pAM2101. The *in vivo* activities of several of these sesquiterpene synthases were confirmed by GC analysis of host cells at 72 hours in shake flasks (FIG. 4), showing that FPP toxicity based growth selection in *Escherichia coli* can be used to screen sesquiterpene synthases for *in vivo* enzyme activity.

Table 12 – Growth of *Escherichia coli* Cells Transformed with Plasmids Comprising Various Sesquiterpene Synthase Coding Sequences

Plasmid	FS coding sequences	Colonies	Incubation Time Until Colonies
pAM1668	None	None	n/a
pAM2096	FS_Cj_Ec	None	n/a
pAM2158	FS_S2D_Sc	None	n/a
pAM2117	FS_Aa_Sc	None	n/a
pAM2104	TDS_Fs_Ec	Yes	1 day
pAM2101	PS_S_Ec	Yes	1 day
pAM1670	FS_S2D_Ec	Yes	1-2 days
pAM2157	FS_Aa_Ec	Yes	> 3 days
pAM2098	FS_Pt_Ec	Yes	> 3 days
pAM2097	FS_Cs_Ec	Yes	> 3 days

Example 5

This example demonstrates the feasibility of using FPP toxicity based growth selection in yeast to screen for sesquiterpene synthases with improved *in vivo* performance.

Strain Y3198 was transformed with 500 ng of Construct A or Construct B, respectively. Host cell transformants were incubated for 6 hours with shaking at 250 rpm in 5 mL YPD medium, and washed twice in 5 mL diH_2O. Half of each washed culture was plated to CSM-L agar plates comprising 2% galactose as a sole carbon source (note that growth on galactose induces expression of the MEV pathway enzyme coding sequences engineered into strain Y3198 resulting in increased FPP production in the cells). The remaining half of each washed culture was plated to CSM-L agar plates comprising 2% glucose as a sole carbon source. Plates were incubated for 3 days at 30°C until individual colonies were approximately 1 mm in diameter.

As shown in FIG. 5, only cells transformed with Construct A were able to grow on CSM-L medium comprising galactose as the sole carbon source, demonstrating the utility of FPP toxicity based growth selection in yeast to screen sesquiterpene synthases for *in vivo* enzyme activity.

Example 6

This example demonstrates the feasibility of using Nile Red fluorescence to screen for terpene synthases with improved *in vivo* performance.

To determine relative farnesene titers, 48 colonies of strain Y3198 transformed with Construct A and 48 colonies of strain Y3198 transformed with expression plasmid pAM404 were picked into separate wells of a 96-well plate containing 360 uL of Bird Seed Medium (BSM) with 2% galactose per well (preculture). After 2 days of incubation at 30°C with 999 rpm agitation, 16 uL of each well was inoculated into a well of a new 96-well plate containing 360 uL of fresh BSM with 2% galactose (production culture). After another 2 days of incubation at 30°C with 999 rpm agitation, samples were taken for farnesene titer determination by Nile Red fluorescence.

For Nile Red fluorescence analysis, 98 uL of each culture was transferred into a 96-well black polystyrene flat bottom assay plate, and 2 uL of Nile Red (Invitrogen, Carlsbad, CA) dissolved at 100 ug/mL in DMSO was added to each well. Fluorescence levels were immediately measured with excitation at 500 nm and emission at 550 nm.

As shown in FIG. 6, the single chromosomally integrated copy of the FS_Aa_Sc coding sequence in strain Y3198 transformed with Construct A produced a Nile Red fluorescence signal that was at 39% of that obtained with strain Y3198 transformed with high-copy expression plasmid pAM404. The approximately 3-fold difference in farnesene titers between the two strains was confirmed by GC analysis (data not shown), demonstrating that there is sufficient difference (delta) in Nile Red fluorescence levels (*i.e.*, farnesene levels) for this system to be a suitable screen for terpene synthases with improved *in vivo* performance.

Example 7

This example demonstrates the feasibility of using sesquiterpene synthase competition in yeast to rank sesquiterpene synthases according to their *in vivo* enzyme activity levels.

For each of yeast strains Y3353 and Y3354, eight single colonies of the original transformation of strain Y3125 with expression plasmid pAM1812 or pAM1813, respectively, were incubated in separate wells of a 96-well plate containing 360 uL of Bird Seed Medium (BSM) with 2% sucrose per well (preculture). After 2 days of incubation at 30°C with 999 rpm agitation, 16 uL of each well was inoculated into a well of a new 96-well plate containing 360 uL of fresh BSM with 4% sucrose (production culture). After another 2 days of incubation at 30°C with 999 rpm agitation, samples were taken and analyzed for terpene production by gas chromatography (GC) analysis. For each strain, a single colony was also restreaked on CSM-L-M-U agar plates, and eight single colonies from each re-streak were grown as described, and analyzed for terpene production by GC analysis.

For GC analysis, samples were extracted with methanol-heptane (1:1 v/v), and the mixtures were centrifuged to remove cellular material. An aliquot of the methanol-heptane extract was diluted into heptane, and then injected onto a methyl silicone stationary phase using a pulsed split injection. Farnesene and trichodiene were separated by boiling point using GC with flame ionization detection (FID). Trans-β-caryophyllene was used as a retention time marker to monitor successful injection and elution during the specified GC oven profile. The titers of farnesene and trichodiene were used to calculate the farnesene/trichodiene ratios.

As shown in Table 13, terpene titers among independent original transformants varied considerably from well to well, resulting in coefficients of variation (CVs) of up to

29%. CVs were reduced for the eight replicates. The observed CVs show that well-to-well and clone-to-clone variations make it impossible to draw conclusions directly from sesquiterpene titers as to the activities of the sequiterpene synthases that produced the sesquiterpenes. However, as also shown in Table 13, farnesene/trichodiene ratios were rather consistent across samples, producing CVs of no more than 5.7%. Thus, by co-expressing a test sesquiterpene synthase from the same plasmid as a comparison sesquiterpene synthase, and by comparing the activities of the test and comparison sesequiterpene synthases, well-to-well and clone-to-clone variations were greatly reduced, making it possible to benchmark the *in vivo* catalytic efficiency of a test sesquiterpene synthase (*e.g.*, a farnesene synthase) against that of a comparison sequiterpene synthase (*e.g.*, a TDS).

Table 13 – β–Farnesene and Trichodiene Titers and Titer Ratios for Yeast Strains Expressing a FS and a TDS from the Same Plasmid

Strain	Coding Sequences (promoters)	Farnesene (mg/L) (CV%)	Trichodiene (mg/L) (CV%)	Farnesene/Trichodiene (CV%)
Y3353	FS_Aa_Sc (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	953 (5.6%)	695 (6.3%)	1.37 (3.0%)
Y3353 restreak		829 (7.1%)	613 (8.5%)	1.36 (5.7%)
Y3354	FS_Aa_Sc (P_{GAL10}) TDS_Fs_Sc (P_{GAL1})	211 (28.9%)	1033 (29.3%)	0.21 (4.3%)
Y3354 restreak		200 (11.2%)	973 (7.9%)	0.21 (4.8%)

To further validate sesquiterpene synthase competition as a ranking tool for sesquiterpene synthase activities in yeast, yeast strains Y3353, Y3394, and Y3395 were evaluated as described. As shown in Table 14, absolute farnesene titers again showed significant variations whereas low CVs were observed for the farnesene/trichodiene ratios. As judged by the observed ratios, the tested farnesene synthases could be ranked as follows (from most active to least active): *Actinidia deliciosa* farnesene synthase expressed from the FS_Ad_Sc coding sequence (Y3395) > *Artemisia annua* farnesene synthase expressed from the FS_Aa_Sc coding sequence (Y3353) > S2D mutant *Artemisia annua* farnesene synthase expressed from the FS_S2D_Ec coding sequence (Y3394). This ranking was confirmed by GC analysis using strains that harbored a single copy of the FS_Aa_Sc, FS_S2D_Ec, or FS_Ad_Sc coding sequence (FIG. 7), thus validating the utility of sesquiterpene synthase competition as a means for ranking sesquiterpene synthases according to their *in vivo* enzyme activity levels in yeast.

Table 14 – β–Farnesene and Trichodiene Titers and Titer Ratios for Yeast Strains Expressing a FS and a TDS from the Same Plasmid

Strain	Coding Sequences (promoters)	Farnesene (mg/L) (CV%)	Trichodiene (mg/L) (CV%)	Farnesene/Trichodiene (CV%)
Y3395	FS_Ad_Sc (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	752 (28.3%)	453 (28.5%)	1.66 (3.0%)
Y3353	FS_Aa_Sc (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	814 (6.0%)	616 (5.0%)	1.32 (2.6%)
Y3394 *	FS_S2D_Ec (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	590 (21.5%)	710 (19.7%)	0.82 (6.2%)

* The S2D mutant *Artemisia annua* farnesene synthase expressed from the FS_S2D_Ec coding sequence was least active likely due to sub-optimal codon usage for the FS_S2D_Ec coding sequence in the yeast host cells.

To yet further validate sesquiterpene synthase competition as a ranking tool for sesquiterpene synthase activities in yeast, yeast strains Y5444, Y5445, Y5446, and Y5447 were evaluated as described with the following exceptions: for each strain four colonies instead of eight colonies were analyzed, cultures were incubated not at 30°C but at 34°C, the preculture lasted not 2 days but 3 days, and the production culture was a 10-fold dilution of the preculture. As shown in Table 15 and FIG. 8, a linear relationship was observed between the number of integrated farnesene synthase coding sequences and farnesene to trichodiene ratios, confirming the utility of sesquiterpene synthase competition as a means for ranking sesquiterpene synthases according to their *in vivo* enzyme activity levels in yeast.

Table 15 - β–Farnesene and Trichodiene Titers and Titer Ratios for Yeast Strains Expressing One Integrated Copy of TDS and Multiple Integrated Copies of FS Coding Sequences

Strain	Copies of Integrated FS_Aa_Sc Coding Sequence	Farnesene (mg/L)	Trichodiene (mg/L)	Farnesene/ Trichodiene
Y5444	2 [a)]	2300±62	811±20	2.8±0.038
Y5445	4 [b)]	2721±109	516±19	5.3±0.057
Y5446	6 [c)]	2761±129	387±22	7.1±0.089
Y5447	8 [d)]	2822±53	306±9	9.2±0.114

a) PGAL1_FS_C_7_TADH1 and pGAL10_FS_D_3.5_TCYC1 integrated at ADH5 locus
b) = a) plus pGAL1_FS_B_5.3_TADH1 and pGAL10_FS_A_5.3_TCYC1 integrated at ADH5 locus
c) = b) plus pGAL1_FS_B_5.3_TADH1 and pGAL10_FS_A_5.3_TCYC1 integrated at BIO4 locus
d) = c) plus pGAL1_FS_C_7_TADH1 and pGAL10_FS_D_3.5_TCYC1 integrated at BIO4 locus

Example 8

This example describes methods for generating libraries of sesquiterpene synthases variants.

Several farnesene synthase variant libraries were generated using the FS_S2D_Ec coding sequence as a template. For each library, 250-500 ng of pAM1670 were subjected to error-prone PCR using the GeneMorph® II Random Mutagenesis Kit (Agilent

Technologies, Inc., Santa Clara, CA) according to manufacturer suggested protocols, and using primers LX-268-139-S2D-F (SEQ ID NO: 106) and LX-268-139-S2D-R (SEQ ID NO: 107) in 25 amplication cycles. The PCR products were gel purified and digested sequentially with FastDigest® NdeI and BamHI restriction endonucleases (Fermentas Inc., Burlington, Ontario). Vector pAM1668 was digested to completion using the same two restriction endonucleases, and the linearized vector DNA fragment was treated with calf intestine alkaline phosphatase (CIP) to remove 5' phosphate groups that would permit recircularization. The purified PCR products and the linearized pAM1688 vector were ligated in an insert to vector ratio of 3:1 using T4 DNA ligase, and 2 uL of the ligation reaction mixture was transformed into XL1-Blue electro-competent *Escherichia coli* cells (Agilent Technologies Inc., Santa Clara, CA) according to manufacturer suggested protocols. Host cell transformants were selected on several LB agar plates (100 mm diameter) comprising carbenicillin. To assess the quality of the random mutagenesis library, 48 or 96 single colonies were picked and grown in LB media comprising carbenicillin, plasmid DNA was isolated from each culture using the QIAprep 96 Turbo Miniprep Kit (Qiagen, Valencia, CA), and the plasmid DNA was digested using FastDigest® NdeI and BamHI restriction endonucleases (Fermentas Inc., Burlington, Ontario) to determine that approximately 95% of plasmids contained an insert. The plasmids were also sequenced to determine the mutation frequency, which was found to be an estimated average of 2-6 nucleotide changes per FS coding sequence. The remaining colonies were washed off the agar plates and plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA).

Several farnesene synthase variant libraries were also generated using the FS_Aa_Sc coding sequence as a template. Construct C was essentially identical to Construct A except that prior to final assembly the FS_Aa_Sc PCR product was mutagenized by error-prone PCR using the Mutazyme II ® Kit (Agilent Technologies, Inc., Santa Clara, CA) and primers AM-288-90-CPK1618 (SEQ ID NO: 83) and AM-288-90-CPK1619 (SEQ ID NO: 84) according to manufacturer suggested protocols. Strain Y3198 was transformed with 500 ng of Construct C, and host cell transformants were recovered for 6 hours with shaking at 250 rpm in 5 mL YPD medium before they were plated out on CSM-L with 2% galactose.

Example 9

This example describes methods for screening libraries of sesquiterpene synthase variants by FPP toxicity based growth selection in *Escherichia coli*.

ElectroMAX DH5α-E *Escherichia coli* cells (Invitrogen, Carlsbad, CA) were transformed with 5 ng of expression plasmid pAM765 (encoding enzymes of the MEV pathway) and 5 ng of the FS_S2D_Ec based farnesene synthase variant library plasmids of Example 8. More than 40 co-transformation experiments were performed to generate ~ 2 x 10^5 host cell transformants. Host cell transformants were plated on LB agar plates (100 mm diameter) comprising carbenicillin and chlorophenicol, and incubated at 30°C for 2 days, after which approximately 400 large colonies and a similar number of small colonies were visible on the agar plates. No colonies were observed when ElectroMAX DH5α-E *Escherichia coli* cells (Invitrogen, Carlsbad, CA) were transformed with 5 ng of expression plasmid pAM765 only were plated on LB agar plates comprising carbenicillin and chlorophenicol. Assuming that these colonies all contained farnesene synthase variants that possessed activity equal or greater than the parent farnesene synthase, the hit rate was estimated to be approximately 1%.

Example 10

This example describes methods for screening libraries of terpene synthase variants using Nile Red fluorescence.

The big colonies obtained in the FPP toxicity based growth selection screen of Example 9 were picked individually into 96-well plates containing 5 uL diH_2O per well, and heated at 98°C for 10 minutes. The generated lysates were transformed into XL1-Blue chemical-competent *Escherichia coli* cells (Agilent Technologies Inc., Santa Clara, CA), and host cell transformants were plated on LB agar plates comprising carbenicillin. Note that chloramphenicol was omitted from the selective medium so that expression plasmid pAM765 was lost from the cells. Individual colonies were picked and grown in LB media comprising carbenicillin, and plasmid DNA was isolated using the QIAprep 96 Turbo Miniprep Kit (Qiagen, Valencia, CA). After the big colonies had been isolated from the selection plates, the small colonies were washed off the plates and their plasmids were isolated as a mixture in a similar fashion.

XL1-Blue chemical-competent *Escherichia coli* cells (Agilent Technologies Inc., Santa Clara, CA) were co-transformed with 5 ng of expression plasmid pAM97 (encoding enzymes of the MEV pathway) and 5 ng of the isolated plasmids (obtained from big colonies) or plasmid mixture (obtained from small colonies). Host cell transformants were plated on LB agar plates containing carbenicillin and chlorophenicol, and incubated at 37°C for 24 hours. Individual colonies were inoculated into 96-well plates containing M9-Hepes media comprising carbenicillin and chlorophenicol, and the cultures were incubated

at 30°C for 24 hours (pre-culture). Then, 50 uL of each culture was used to inoculate a second culture at an initial OD of 0.05. To induce the expression of the MEV pathway enzymes and farnesene synthase, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to each culture at 1 mM final concentration. The cultures were incubated for at least 20 hours (production culture) before farnesene titers were determined by Nile Red fluorescence. An *Escherichia coli* strain co-transformed with pAM97 and pAM1419 was used as the negative control, and an *Escherichia coli* strain co-transformed with pAM97 and pAM1421 was used as the positive control. To validate the Nile Red fluorescence analysis, farnesene titers were also determined by GC analysis.

As shown in FIG. 9, measured Nile Red fluorescence signals directly correlated with farnesene titers determined by GC analysis, further validating Nile Red fluorescence as a means of measuring farnesene titers.

The top 70 strains derived from cells transformed with plasmids obtained from large colonies that produced the largest farnesene titers as determined by Nile Red fluorescence and GC analysis were replica-plated into new 96-well plates, and their production levels were re-measured as described. As shown in FIG. 10, approximately 50 of these strains produced higher farnesene titers than the control strain comprising the parent FS_S2D_Ec coding sequence, with the best farnesene synthase variant producing 65% more farnesene than the parent farnesene synthase.

The plasmids extracted from the approximately 400 large colonies were combined in equal molar amounts, the FS coding sequences were PCR amplified using primers LX-268-130-3-S2D-F (SEQ ID NO: 110) and LX-268-130-4-S2D-R (SEQ ID NO: 109), and the PCR products were gel purified. Vector pAM1734 was linearized using FastDigest® Bst1101 restriction endonuclease (Fermentas Inc., Burlington, Ontario), and the linearized vector was cleaned using the Zymo DNA Clean & Concentrator™ Kit (Zymo Research Corp., Orange, CA). The purified vector and PCR products were mixed in a ratio of 1:3 (vector:insert), and transformed into either strain Y539 or strain Y3198 for ligation via homologous recombination. Host cell transformants were plated on CSM-L agar plates with 2% glucose (Y539 host) or 2% galactose (Y3198 host) as the sole carbon source. Approximately 2,500 individual colonies were picked and farnesene titers were determined by Nile Red fluorescence as described (using BSM with 2% glucose for the Y539 host and BSM with 2% galactose for the Y3198 host). Clones that produced a fluorescence signal that was greater than the signal obtained for the parent control (pAM1764 (FS_S2D_Ec on CEN.ARS plasmid) in Y539) by three times the standard

deviation were re-plated onto CSM-L agar plates to obtain single colonies, and four colonies of each restreak were re-tested by Nile Red fluorescence and GC analysis. As shown in FIGs. 11A and 11B, a number of cells transformed with farnesene synthase variant produced higher titers of farnesene than the parent control strain, and a few produced even higher titers than strain Y539 transformed with expression plasmid pAM1765 (FS_Aa_Ec on high-copy plasmid).

Example 11

This example describes methods for screening libraries of sesquiterpene synthase variants by FPP toxicity based growth selection in yeast.

The washed transformation of strain Y3198 transformed with Construct C of Example 8 was plated to CSM-L agar plates comprising 2% galactose as a sole carbon source. Plates were incubated for 5 days at 30°C and the colonies were picked, propagated, and analyzed for farnesene titers by Nile Red fluorescence and GC analysis as described. As shown in FIG. 12, approximately 15% of clones grown on galactose as the sole carbon source had at least 15% higher farnesene titers than the average titer obtained with the parent control (strain Y3198 transformed with Construct A). Clones that had at least 15% higher farnesene titers than the parent control were streaked for individual colonies to CSM-L agar plates comprising 2% glucose. For each clone, eight individual colonies were picked and re-tested by GC analysis as described. Clones that maintained >15% higher average farnesene titer than the control through GC assay were promoted for the sesquiterpene competition assay.

Example 12

This example describes methods for screening sesquiterpene synthase variants by sesquiterpene synthase competition in yeast.

Plasmids were isolated from the top farnesene producing yeast strains of Example 10 using the Zymoprep™ Yeast Plasmid Miniprep Kit (Zymo Research Corp., Orange, CA). The plasmids were digested using BamHI and NheI restriction endonucleases, and the farnesene synthase variant coding sequences were gel purified and ligated with expression vector pAM1812 digested with the same two restriction endonucleases, thus replacing the FS_Aa_Sc coding sequence of pAM1812 with the variant FS_S2D_Ec or FS_Aa_Sc coding sequences. The new plasmids were transformed into strain Y3125.

Alternatively, and in addition, farnesene synthase variant coding sequences were PCR amplified from isolated plasmids or from cell lysates of top farnesene producing

yeast strains of Example 11 using primers AM-288-90-CPK1618 (SEQ ID NO: 83) and AM-288-90-CPK1619 (SEQ ID NO: 84), the PCR products were gel purified using the E.Z.N.A. ® Gel Extraction Kit (Omega Bio-Tek Inc., Norcross, GA) according to manufacturer suggested protocols, and ligated with expression plasmid pAM1948 digested with FastDigest BamHI restriction endonuclease (Fermentas Inc., Burlington, Ontario) via homologous recombination by transforming strain Y3125 with both purified DNA fragments, thus replacing the IS_Pn_Sc coding sequence of pAM1948 with the variant FS_S2D_Ec coding sequences.

Host cell transformants were plated to CSM-L agar plates with 2% glucose as a sole carbon source, and then replica-plated to CSM-L-U agar plates at 30°C for at least 72 hours to loop out the GAL80 coding sequence in the host cell transformants. For each clone, eight colonies were picked, propagated, and their terpene titers were determined by GC analysis as described. Farnesene synthase variants that produced a higher farnesene/trichodiene ratio than the ratio obtained for the parent farnesene synthase (strain Y3125 transformed with pAM1812) were re-streaked to CSM-L agar plates. For each re-streak, eight individual colonies were re-tested in a 96-well plate production experiment as described. A total of 11 yeast strains comprising farnesene synthase variant showed improved farnesene/trichodiene ratios, suggesting that these farnesene synthase variants possessed improved *in vivo* enzyme activity in yeast.

Example 13

This example describes the characterization of sesquiterpene synthase variants that have improved enzyme activity.

To identify possible causal mutations, the improved farnesene synthase variants identified in Example 12, and in other libraries of farnesene synthase variants generated and screened essentially as described, were sequenced by Elim Biopharmeceuticals, Inc. (Hayward, CA). The identified mutations are listed in Table 16. Overall, an estimated total of 300,000 clones were screened and evaluated by FPP toxicity based growth selection in *Escherichia coli* and/or yeast, Nile Red fluorescence and/or GC analysis, and sesquiterpene synthase competition in yeast, leading to the identification of 51 farnesene synthase variants with improved enzyme activity. A number of farnesene synthase variants were identified more than once (number of times are indicated in Table 16 in parenthesis next to the variant names), and certain nucleotide and amino acid changes were found in multiple farnesene synthase variants, suggesting causal relationships between the mutations and the increased enzyme activity. In some farnesene synthase variants only

silent mutations were uncovered, suggesting that these mutations likely improved expression of the encoded FS.

Table 16 – Mutations identified in farnesene synthase variants with improved enzyme activity

Variant FS	Nucleotide change	Amino acid change	Farnesene/trichodiene ratios over parent FS
1	1343 (A→G)	I434T	163%
2	71 (T→A) 182 (T→A) 271 (C→T) 955 (A→T)	V24D L61Q silent T319S	150%
3	71 (T→A) 587 (C→G) 1092 (T→C) 1267 (G→A) 1399 (G→A)	V24D T196S silent V423I V467I	136%
4	441 (T→C) 807 (A→G) 839 (T→A) 1301 (T→C) 1599 (G→A) 1629 (T→C)	silent silent L280Q I434T silent silent	129%
5	343 (A→G) 587 (C→G) 955 (A→T)	I115V T196S T319S	129%
6	429 (T→C) 751 (C→A) 861 (G→T) 863 (A→T) 921(A→G) 957 (C→A) 1325 (G→A)	silent L251M silent Y288F silent silent G442D	126%
7	168 (A→T) 215 (A→T) 431 (T→A) 1325 (G→C) 1378 (A→G)	silent E72V F144Y G442A I460V	122%
8	32 (T→C) 225 (T→C) 345 (A→G) 1715 (T→A)	F11S Silent I115M M572K	181%
9	32 (T→C) 1015 (T→C) 1605 (A→G)	F11S Silent Silent	180%
10	598 (T→C) 631 (T→A) 1105 (G→C)	Silent S211T V369L	156
11 (7)	11 (T→C)	L4S	154%
12 (2)	8 (C→A)	T3N	175%
13 (3)	8 (C→A) 870 (T→C)	T3N Silent	188%
14	867 (A→G)	Silent	124%

	1587 (C→A)	Silent	
15	867 (A→G) 1591 (C→T)	Silent P531S	127%
16	867 (A→G)	Silent	116%
17	17 (T→C)	I6T	149%
18	315 (A→T) 867 (A→G)	E105D Silent	124%
19	1337 (C→A)	T446N	125%
20	1667 (C→T)	A556V	130%
21	238 (A→G)	N80D	154%
22	1056 (A→G)	Silent	163%
23	1098 (A→G)	Silent	156%
24	1111 (T→A)	L371M	156%
25	1542 (C→A)	Silent	158%
26	627 (C→T)	Silent	129%
27	1077 (A→T)	E359T	128%
28	1153 (A→G) 1463 (C→T)	T385A S488F	132%
29	104 (T→C) 1584 (T→C)	M35T Silent	125%
30	1245 (T→C) 1299 (G→A) 1509 (A→T)	Silent M433I Silent	137%
31	265 (A→G) 1192 (T→G) 1330 (A→C) 1371 (T→C) 1533 (A→G) 1636 (T→C) 1701 (T→C)	I89V l398v I444L silent Silent Silent Silent	127%
32	519 (A→G)	silent	120%
33	1491 (T→C)	silent	119%
34	1514 (A→T)	E505V	115%
35	1576 (A→T)	T526S	125%
36	59 (T→A)	V20E	128%
37 (3)	214 (G→A)	E72K	128%
38	26 (T→A) 39 (A→G) 113 (A→G) 148 (G→A) 549 (T→A) 1723 (T→A)	V9D silent N38S D50N silent stop575K	129%
39	792 (A→G)	silent	133%
40	1100 (A→T) 1484 (A→G) 1542 (C→T)	E357V E495G silent	139%

Example 14

This example describes methods for combining mutations and screening for improved farnesene synthase variants using sesquiterpene synthase competition in yeast.

Various amino acid changes identified in Example 13 were combined by sequence overlap extension (SOE; *Ho, et al, 1989*), and the combinations were screened by

sequiterpene synthase competition as described to identify FS with increased enzyme activity. As shown in Table 17, several farnesene synthase variants showed substantial improvements in their farnesene/trichodiene ratios over that of the wild-type *A. annua* FS. The increased activity of several of these farnesene synthase variants was further confirmed by GC analysis of yeast strains comprising single chromosomally integrated copies of the farnesene synthase variants coding sequences (FIG. 14).

Table 17 – Variant FS by combination of mutants

Variant FS	Parent FS coding sequence	Amino Acid mutations relative to wild-type *A. annua* FS	Fold increase in enzyme activity over activity of WT FS as measured by farnesene/trichodiene ratios
G6	FS_A (SEQ ID NO: 118)	F11S, V24D, I115M, V423I, G442A, I460V, V467I	3
FS_D_3.5 (SEQ ID NO: 108)	FS_D	F11S, V24D, M35T, T196S, Y288F, T319S, I434T, T446N, I460V, V467I	3.5
FS_A_4A	FS_A (SEQ ID NO: 118)	F11S, V24D, I115M, S211T, V369L, V423I, I434T, G442A, I460V, V467I	4
FS_A_4B	FS_A (SEQ ID NO: 118)	F11S, V24D, I115M, T196S, T319S, V423I, I434T, G442A, I460V, V467I	4
FS_A_4.5	FS_A (SEQ ID NO: 118)	F11S, V24D, I115M, T196S, Y288F, T319S, V423I, I434T, G442A, I460V, V467I	4.3
FS_A_5.3	FS_A (SEQ ID NO: 118)	F11S, V24D, M35T, I115M, T196S, Y288F, T319S, I434T, T446N, I460V, V467I	5.3
FS_B_5.3	FS_B (SEQ ID NO: 119)	F11S, V24D, M35T, T196S, Y288F, T319S, I434T, T446N, I460V, V467I	5.3
FS_A_6	FS_A (SEQ ID NO: 118)	F11S, V20E, V24D, M35T, I115M, T196S, Y288F, T319S, T385A, I434T, T446N, I460V, V467I	6
FS_C_6	FS_C (SEQ ID NO: 120)	F11S, V24D, S211T, V369L, V423I, I434T, G442A, I460V, V467I	6
FS_C_7	FS_C (SEQ ID NO: 120)	F11S, V24D, M35T, T196S, Y288F, T319S, I434T, T446N, I460V, V467I	7
FS_C_8	FS_C (SEQ ID NO: 120)	F11S, V24D, L18I, M35T, T196S, Y288F, T319S, R348K, T385A, I434T, T446N, I460V, V467I	8

Example 15

This example describes methods for expressing and purifying wild-type farnesene synthase and farnesene synthase variant proteins.

Escherichia coli Rosetta (DE3) cells were transformed with expression plasmids comprising the FS_Aa_Ec, FS_B_5.3_Ec (FS_B_5.3 codon-optimized for expression in *Escherichia coli*), or FS_C_8 coding sequence cloned into the BamHI and NdeI sites of vector pAM1490 (SEQ ID NO: 117). Expression of FS in 1 L cultures of host cell transformants were induced by adding 0.4 mM IPTG, and each culture was incubated at 20°C for 20 hours. Cell lysates were loaded onto His GraviTrap columns pre-packed with Ni Sepharose™ 6 Fast Flow resin (GE Healthcare, Piscataway, NJ). The columns were washed with 10 mL (10 column volumes) Binding Buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole, 5% glycerol, 0.5 mM DTT), and fractions were step-wise eluted using 4 ml Elution Buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 0.5 mM DTT) comprising 50 (Fraction 1), 100 (Fraction 2), or 250 mM (Fraction 3) imidazole. All fractions were dialyzed against Binding Buffer, and Fraction 2 was further dialyzed against Protease Digestion Buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol, 1 mM DTT) in a dialysis cassette. To remove the His_6-tags, 120 U of PreScission Protease (GE Healthcare, Piscataway, NJ) was added directly to each dialysis cassette, and dialysis was allowed to continue overnight before the protease-digested samples were dialyzed against Storage Buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 20% glycerol, 1 mM DTT) and were removed from the dialysis cassette. To remove the GST-tagged protease, Glutathione Sepharose 4B beads (GE Healthcare, Piscataway, NJ) pre-washed with the Storage Buffer were added to the protease-digested samples, and the bead mixtures were incubated for one hour with gentle mixing. The FS proteins were finally recovered by passing the mixtures through a Poly-Prep Chromatography Column (Bio-Rad, Hercules, CA), and the purified FS proteins were stored at -80 °C.

The isolated farnesene synthase variants were assayed for kinetic properties. As shown in Table 18, the increased *in vivo* activity of the farnesene synthase variants was reflected in an increased k_{cat} of the isolated farnesene synthase variants.

Table 18 – Steady-state kinetic parameters of wild-type and farnesene synthase variants

FS coding sequence	k_{cat} (s^{-1})	Relative k_{cat}	K_m (uM)
FS_Aa_Ec	0.071 ± 0.004	1	0.58 ± 0.19
FS_B_5.3_Ec	0.093 ± 0.009	1.31	1.18 ± 0.44
FS_C_8	0.139 ± 0.012	1.96	1.55 ± 0.76

Example 16

This example describes the use of sesquiterpene synthase competition to select suitable promoters to tune farnesene synthase expression to a level that enables FPP toxicity-based growth selection in yeast.

Various promoters were PCR amplified as described in Table 19.

Table 19 – Templates and primers used to PCR amplify *S. cerevisiae* promoters

Promoter of Gene	Template	PCR primer 1	PCR primer 2
FBA1	100ng of Y002 genomic DNA	AM-288-160-CPK2063 (SEQ ID NO: 141)	AM-288-160-CPK2064 (SEQ ID NO: 142)
TPI1		AM-288-160-CPK2061 (SEQ ID NO: 139)	AM-288-160-CPK2062 (SEQ ID NO: 140)
HSP12		AM-288-160-CPK2065 (SEQ ID NO: 143)	AM-288-160-CPK2066 (SEQ ID NO: 144)
PET9		AM-288-160-CPK2085 (SEQ ID NO: 151)	AM-288-160-CPK2086 (SEQ ID NO: 152)
QCR9		AM-288-160-CPK2091 (SEQ ID NO: 114)	AM-288-160-CPK2092 (SEQ ID NO: 115)
QCR7		AM-288-160-CPK2087 (SEQ ID NO: 153)	AM-288-160-CPK2088 (SEQ ID NO: 154)
CDC19		AM-288-160-CPK2055 (SEQ ID NO: 135)	AM-288-160-CPK2056 (SEQ ID NO: 136)
TEF2		AM-288-160-CPK2073 (SEQ ID NO: 147)	AM-288-160-CPK2074 (SEQ ID NO: 148)
PGK1		AM-288-160-CPK2079 (SEQ ID NO: 149)	AM-288-160-CPK2080 (SEQ ID NO: 150)
SIP18		AM-288-160-CPK2067 (SEQ ID NO: 145)	AM-288-160-CPK2068 (SEQ ID NO: 146)
ERG8		AM-288-160-CPK2051 (SEQ ID NO: 131)	AM-288-160-CPK2052 (SEQ ID NO: 132)
GRE1		AM-288-160-CPK2053 (SEQ ID NO: 133)	AM-288-160-CPK2054 (SEQ ID NO: 134)
PDC1		AM-288-160-CPK2059 (SEQ ID NO: 137)	AM-288-160-CPK2060 (SEQ ID NO: 138)
ANB1		AM-288-160-CPK2089 (SEQ ID NO: 155)	AM-288-160-CPK2090 (SEQ ID NO: 113)

PCR amplifications were done using the Phusion High Fidelity DNA Polymerase System (Finnzyme Inc, Espoo, Finland). PCR products were gel purified using the E.Z.N.A. ® Gel Extraction Kit (Omega Bio-Tek Inc., Norcross, GA) according to manufacturer's suggested protocols.

To replace the GAL1 promoter in pAM2191, the expression plasmid was digested using FastDigest® BamHI restriction endonuclease, and 100 ng of this linearized plasmid was co-transformed with 300 ng of each of the promoters PCR products into exponentially growing Y3198 cells. Host cell transformants were plated to CSM-L with 2% glucose as a sole carbon source, and incubated for 3 days at 30°C until individual colonies were ~1mm in diameter. Eight colonies from each transformation were picked and grown as described, before their farnesene and trichodiene titers were determined by GC analysis as described (Table 20). Promoters that gave farnesene/trichodiene ratios similar to that of the GAL1

promoter linked to the FS_Aa_Sc coding sequence were selected and further screened by FPP-toxicity based growth selection in Y227 as described. Of the tested promoters, the promoter of the PET9 gene was shown to be suitable for use in FPP toxicity-based selection for improved FS when using the FS_A_5.3 coding sequence as the parent template.

Table 20 – Farnesene to trichodiene titers obtained using various promoters to drive expression of the FS_A_5.3 coding sequence

Promoter of Gene	FS coding sequence	Farnesene/Trichodiene	% Strength of GAL1 promoter
GAL1	FS_Aa_Sc	1.06	1
FBA1	FS_A_5.3	1.18	0.222
TPI1		0.875	0.165
HSP12		0.87	0.164
PET9		0.71	0.133
QCR9		0.6	0.113
QCR7		0.525	0.099
CDC19		0.19	0.035
TEF2		0.19	0.035
PGK1		0.19	0.035
SIP18		0.16	0.030
ERG8		0.137	0.025
GRE1		0.093	0.017
PDC1		0.085	0.016
ANB1		0.072	0.013

Example 17

This example describes methods for porting beneficial mutations of farnesene synthase, identified through sesquiterpene synthase competition, into a related sesquiterpene synthase.

Various amino acid changes of FS identified in Example 13 were confirmed as causal for improvement in FS activity using saturation mutagenesis, and several of these mutations were ported into a related sesquiterpene synthase, amorphadiene synthase of *Artemisia annua* (ADS), to determine whether a similar improvement in activity could be achieved. The amino acid positions of FS selected for porting were Met 35, Tyr288, Thr 319, Val 369, Ile 434, Thr 446, I460 and V467. Based on multiple sequence alignments,

the aligned corresponding amino acids in ADS are Ala13, Cys260, Ala291, Met341, Thr406, Thr418, Phe432 and Gly439. Each of these positions were mutated by substitution with 19 other residues, and each mutant was tested in a amorphadiene:trichodiene competition assay.

The ADS mutants were constructed by using separate PCR reactions to amplify two DNA fragments that overlap at a specific codon in the ADS gene. Each oligonucleotide pair was synthesized with a mismatched degenerate nucleotide sequence in the middle of the oligonucleotide (for example, NNK, wherein K represents mixed G and T bases, and N represents A, T, G and C mixed bases), flanked on both sides by nucleotides that specifically anneal to the target region.

Table 21 – Primers used in Constructing ADS-TDS 2μ plasmid

Primer	**Sequence (5' to 3')**
ADS-A13-F	GGAGAAAAAACCCCGGATCCATGGCCTTGACTGAAGAGAAA CCTATAAGGCCAATTNNKAATTTCCCACCTTCTATTTG (SEQ ID NO: 157)
ADS-C260-F	CATTTGATATTAAGAAAAATGCCCCANNKCTGAGAGATCGTA TCGTTGAATGC (SEQ ID NO: 158)
ADS-C260-R	CATTTTTCTTAATATCAAATGCCTTCCACC (SEQ ID NO: 159)
ADS-A291-F	CAAGAGCTAGAGTTTTTTTCACTAAGNNKGTTGCTGTGATAA CACTTATTG (SEQ ID NO: 160)
ADS-A291-R	GAAAAAAACTCTAGCTCTTGAATAC (SEQ ID NO: 161)
ADS-M341-F	CATGAAGCCTATTTACAAATTATTCNNKGATACCTACACAGA AATG (SEQ ID NO: 162)
ADS-M341-R	AATTTGTAAATAGGCTTCATGTATTC (SEQ ID NO: 163)
ADS-M406-F	GTTATTATCACTGGTGGTGCAAACTTGCTANNKACTACCTGT TATCTAG (SEQ ID NO: 164)
ADS-M406-R	GTTTGCACCACCAGTGATAATAACTACGGG (SEQ ID NO: 165)
ADS-M418-F	GTTATCTAGGAATGAGCGATATTTTCNNKAAGGAATCAGTTG AGTGGGC (SEQ ID NO: 166)
ADS-M418-R	CGCTCATTCCTAGATAACAGGTAG (SEQ ID NO: 167)
ADS-F432-F	GGGCTGTATCTGCTCCGCCTTTANNKCGTTACAGTGGTATTCT G (SEQ ID NO: 168)
ADS-F432-R	GGCGGAGCAGATACAGCCCACTCAAC (SEQ ID NO: 169)
ADS-G439-F	CTTTATTCCGTTACAGTGGTATTCTGNNKAGGAGATTAAATG ACCTGATG (SEQ ID NO: 170)
ADS-G439-R	GAATACCACTGTAACGGAATAAAG (SEQ ID NO: 171)
ADS-SM-5'	GAAAAAACCCCGGATCCATGGCCTTGACTGAAGAGAAAC (SEQ ID NO: 172)
ADS-SM-3'	GGTTAGAGCGGATCTTAGCTAGCTTAGATAGACATAGGGTAA AC (SEQ ID NO: 173)

Table 22. List of PCR reactions for generating ADS saturation mutants

AA position	Fwd oligonucleotide	Rev oligonucleotide	Fwd oligonucleotide	Rev oligonucleotide
Ala13	ADS-A13-F	ADS-SM-3'		
Cys260	ADS-SM-5'	ADS-C260-R	ADS-C260-F	ADS-SM-3'
Ala291	ADS-SM-5'	ADS-A291-R	ADS-A291-F	ADS-SM-3'
Met341	ADS-SM-5'	ADS-M341-R	ADS-M341-F	ADS-SM-3'
Thr406	ADS-SM-5'	ADS-M406-R	ADS-M406-F	ADS-SM-3'
Thr418	ADS-SM-5'	ADS-M418-R	ADS-M418-F	ADS-SM-3'
Phe432	ADS-SM-5'	ADS-M432-R	ADS-M432-F	ADS-SM-3'
Gly439	ADS-SM-5'	ADS-M439-R	ADS-M439-F	ADS-SM-3'

To prepare ADS-TDS competition vectors, pAM1948 plasmid was digested with BamHI and NheI to excise the IS_Pn_Sc coding sequence, and gel purified using the Zymoclean Gel purification kit (Zymo Research, Irvine CA). Amplification of the ADS open reading frame (orf) containing each saturation mutant was performed by mixing equimolar amounts of a PCR fragment listed in Table 22 and the appropriate ADS-SM-5' and ADS-SM-3' oligos listed in Table 21, with the exception that for mutagenesis of residue Ala13, ADS-A13-F and ADS-SM-3' were used to amplify the orf directly. The amplified PCR products were gap-repaired in linearized pAM1948 and transformed in Y3125. For each site saturation mutant, the fold-improvement was obtained by comparing the ADS/TDS ratio of the mutant to that of the ADS WT.

As shown in FIG. 17, substitution of Ala291 with either a valine (A291V; SEQ ID NO: 174), cysteine (A291C; SEQ ID NO: 175) or a isoleucine (A291I; SEQ ID NO: 176) resulted in improvement of the ADS/TDS ratio by more than 30% compared to the ADS parent (WT).

To confirm the improved activities of ADS A291V, A291C and A291I, amorphadiene titers were determined in cells transformed with expression plasmids for each mutant. The ADS-TDS plasmid containing A291V, A291C and A291I, respectively, was digested using BamHI and NheI and gel extracted. The CEN.ARS plasmid and 2μ Leu2 plasmid were linearized using the same restriction enzyme and gel purified. The digested fragments containing ADS mutants were ligated into linearized CEN.ARS or 2μ Leu2 plasmid using T4 ligase at 16 °C. 2 μl of the reaction was transformed into Xl1-Blue cells and plated on LB plates. Colonies containing each ADS mutant either in CEN.ARS plasmid or 2μ Leu2 plasmid were sequence verified.

Table 23. List of plasmids encoding ADS variants	
Plasmid No.	Description
pAM2507	*Artemesia Annua* (Aa) ADS mutant A291C on CEN.ARS plasmid using pAM1734 as the backbone
pAM2506	Aa.ADS mutant A291I on CEN.ARS plasmid using pAM1734 as the backbone
pAM2505	Aa.ADS mutant A291V on CEN.ARS plasmid using pAM1734 as the backbone
pAM2504	Aa.ADS WT on CEN.ARS plasmid using pAM1734 as the backbone
pAM2503	PGAL1-Aa.ADS mutant A291C in leu2 2u
pAM2502	PGAL1-Aa.ADS mutant A291I in leu2 2u
pAM2501	PGAL1-Aa.ADS mutant A291V in leu2 2u
pAM2419	Aa.ADS mutant A291I (under pGAL1) and TDS (under PGAL10) in leu2d
pAM2418	Aa.ADS mutant A291C (under pGAL1) and TDS (under PGAL10) in leu2d
pAM2417	Artemesia Annua ADS mutant A291V (under pGAL1) and TDS (under PGAL10) in leu2d
pAM2414	Aa. ADS (under pGAL1) and TDS (under PGAL10) in leu2d

For production, each plasmid was transformed into cured Y227 and plated on CSM-L plates. Eight colonies were picked for each mutant and grown in a 96-well plate containing 360 µl of Bird Seed Media with 4% galactose. After 2 days of incubation at 34 °C, 16 µl of each well was inoculated into a new 96-well plate containing fresh Bird Seed Media with 4% galactose for production. After 2 days of incubation at 30 °C, production samples were taken for Nile Red and GC analysis.

As shown in Figure 18, each of ADS A291V, A291C and A291I showed an increase in amorphadiene production compared to parent ADS when expressed on either a 2µ plasmid or CEN.ARS plasmid. The best mutant, Ala291Val, showed up to 58% improvement in amorphadiene titer than that of the parent.

These results demonstrate that beneficial mutations identified in one terpene synthase using the terpene synthase competition assay can be successfully ported into a related terpene synthase to effect improved synthase activity.

Example 18

This example demonstrates the feasibility of using sesquiterpene synthase competition in yeast to rank patchoulol synthases (PS) according to their *in vivo* enzyme activity levels.

Yeast strains Y9259, Y11136, Y9260 were made by transforming expression plasmids pAM2596, pAM2702, and pAM2597 into yeast strain Y9120 (comprising the MEV pathway), respectively, and confirmed by colony PCR to contain the correct plasmid. Each plasmid contains a different PS isoform on a GAL1 promoter and identical

versions of trichodiene synthase (TDS) on the divergent GAL10 promoter. The confirmed colony was re-streaked for single colonies, from which eight colonies were incubated in separate wells of a 96-well plate containing 360 uL Bird Seed Medium (BSM) with 2% sucrose per well (preculture). After 2 days of incubation at 30°C with 999 rpm agitation, 6.4 uL of each well was inoculated into a well of a new 96-well plate containing 150 uL of fresh BSM with 4% galactose and 3.33% mineral oil and Brij-56 emulsion (production culture). After another 4 days of incubation at 30°C with 999 rpm agitation, samples were taken and analyzed for terpene production by gas chromatography (GC) analysis.

For GC analysis, samples were extracted with methanol-butoxy ethanol-heptane (100uL:50uL:400uL v/v), and the cell material was allowed to settle by gravity. An aliquot of the heptane extract was further diluted into heptane, and then injected onto a methyl silicone stationary phase using a pulsed split injection. Patchouli alcohol and trichodiene were separated by boiling point using GC with flame ionization detection (FID). Hexadecane was used as a retention time marker to monitor successful injection and elution during the specified GC oven profile. The titers of patchouli alcohol and trichodiene were used to calculate the patchouli oil/trichodiene ratios (patchouli oil is approximately three times the titer of patchouli alcohol).

As shown in Table 24, patchouli oil/trichodiene ratios were fairly consistent across samples, producing CVs of no more than 5.47%. Thus, by co-expressing PS from the same plasmid as a TDS, the tested patchoulol synthases could be ranked as follows (from most active to least active): PS_isoform_3 coding sequence (Y9260) > PS_isoform_2 coding sequence (Y11136) > PS_isoform_1 coding sequence (Y9259).

Table 24 – Patchouli Oil and Trichodiene Titers and Titer Ratios for Yeast Strains Expressing a PS and a TDS from the Same Plasmid

Strain	Coding Sequences (promoters)	Patchouli Oil/Trichodiene (CV%)
Y9259	PS_isoform_1 (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	0.17 (2.16%)
Y11136	PS_isoform_2 (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	0.68 (5.47%)
Y9260	PS_isoform_3 (P_{GAL1}) TDS_Fs_Sc (P_{GAL10})	1.24 (1.82%)

Example 19

This example demonstrates the feasibility of using monoterpene synthase competition to rank monoterpene synthases according to their *in vivo* enzymatic activity.

To determine the productivity of synthases for the monoterpene limonene, a competition vector was prepared comprising the coding sequence for myrcene synthase of *Ocimum basilicum*, used here as the comparison terpene synthase, being driven by a pGal10 promoter, and the coding sequence for a query limonene synthase (LS) being driven by pGAL1 on the same plasmid (pAM2645). Eight competition plasmids each encoding a different LS truncation variant or isoform were transformed into yeast strain Y8270 (comprising the MEV pathway), respectively, and transformed strains were confirmed by colony PCR to contain the correct plasmid. Eight single colonies of each strain were incubated in 96-well microtiter plates with 360 µL of Bird Seed Medium containing 2% glucose as a carbon source. After 3 days of growth at 30°C with shaking at 998 RPM, 6 µL of each culture was inoculated into a well of a 96-well plate containing 75 µL of Bird Seed Medium containing 4% galactose and 75 µL of isopropyl myristate per well. The production plates were sealed with a Velocity 11 heat sealer (Agilent Technologies) and grown at 30°C with shaking at 998 RPM prior to flash freezing for 2 hours at -20°C. The products were then assayed after quickly adding 300 µl of ethyl acetate containing .001% hexadecane internal standard, heat sealing, and shaking for 2 hours at room temperature. Gas chromatogram flame ionization detection (GC-FID) was performed using standard curves of absolute concentrations of myrcene and limonene. In detail, 2 µL of the ethyl acetate extract was injected onto a methyl silicone stationary phase column with a split ratio of 1:50. This injection was analyzed using hexadecane as an internal standard for injection accuracy and retention time adjustment. Oven temperatures were ramped from 25°C-250°C over the course of 2.5 minutes, at which point the oven was rapidly cooled for the next sample. The titers of myrcene and the limonene were used to calculate the competition ratio.

As shown in FIG. 18, the terpene synthase competition assay can be used to rank the relative performance of different limonene synthase truncation variants and isoforms.

Various modifications and variations of the present disclosure will be apparent to those skilled in the art without departing from the scope and spirit of the disclosure. Although the disclosure has been described in connection with specific preferred embodiments, it should be understood that the claims should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the disclosure, which are understood by those skilled in the art are intended to be within the scope of the claims.

International Patent Application No. PCT/US2012/023446
Appendix B: AMENDED CLAIMS (clean version)

WHAT IS CLAIMED IS:

1. A method of testing a terpene synthase variant for improved *in vivo* performance, comprising the steps of:

(a) dividing a population of host cells into a control population and a test population;

(b) expressing in the control population a control terpene synthase and a comparison terpene synthase, wherein the control terpene synthase can convert a polyprenyl diphosphate to a first terpene, and wherein the comparison terpene synthase can convert the polyprenyl diphosphate to a second terpene;

(c) expressing in the test population a test terpene synthase and the comparison terpene synthase, wherein the test terpene synthase is a variant of the control terpene synthase, and wherein the comparison terpene synthase is expressed at similar levels in the test population and in the control population; and

(d) measuring the amount of the first terpene and the amount of the second terpene in both the test population and in the control population.

2. The method of claim 1, wherein the test population is identified as comprising a terpene synthase variant having improved *in vivo* performance compared to the control terpene synthase by: (i) an increase in the ratio of the amount of the first terpene over the amount of the second terpene in the test population compared to that in the control population; or (ii) a decrease in the ratio of the amount of the second terpene over the amount of the first terpene in the test population compared to that in the control population.

3. The method of claim 1, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are sesquiterpene synthases, and the polyprenyl diphosphate is farnesyl diphosphate (FPP).

4. The method of claim 3, wherein the comparison terpene synthase is a trichodiene synthase.

5. The method of claim 3, wherein the control and test terpene synthases are synthases having farnesene synthase activity.

6. The method of claim 1, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are monoterpene synthases, and the polyprenyl diphosphate is geranyl diphosphate (GPP).

7. The method of claim 1, wherein the host cell comprises one or more heterologous enzymes of the MEV pathway.

8. The method of claim 1, wherein the one or more heterologous enzymes of the MEV pathway are selected from the group consisting of ERG8, ERG10, ERG12, ERG13, ERG19, and HMG1.

International Patent Application No. PCT/US2012/023446
Appendix B: AMENDED CLAIMS (clean version)

9. The method of claim 1, wherein the test terpene synthase differs from the control terpene synthase by 1 to 10 amino acids.

10. The method of claim 2, wherein the ratio of the first terpene over the second terpene is at least 1.3.

11. The method of claim 1, wherein test terpene synthase is expressed at a similar level in the test population as the control terpene synthase is expressed in the control population.

12. The method of claim 1, wherein the control terpene synthase and the test terpene synthase are identical but are expressed at different levels in the control population and the test population.

13. The method of claim 1, wherein the population of host cells is a population of yeast cells.

14. A composition comprising two cell subpopulations derived from a common population of host cells, wherein:

(a) the first subpopulation comprises a control terpene synthase and a comparison terpene synthase, wherein the control terpene synthase converts a polyprenyl diphosphate to a first terpene, and wherein the comparison sesquiterpene synthase converts the polyprenyl diphosphate to a second terpene; and

(b) the second subpopulation comprises a test terpene synthase and the comparison terpene synthase, wherein the test terpene synthase converts the polyprenyl diphosphate to the first terpene, wherein the test terpene synthase is a variant of the control terpene synthase, and wherein the comparison terpene synthase is expressed at similar levels in the first subpopulation and in the second subpopulation.

15. The composition of claim 14, wherein the ratio of the first terpene over the second terpene is greater in the second subpopulation compared to that in the first subpopulation.

16. The method of claim 1, wherein the population of host cells is a population of *E. coli* cells.

17. The method of claim 8, wherein the host cell comprises a heterologous isopentyl pyrophosphate (IPP) isomerase.

18. The method of claim 8, wherein the host cell comprises a heterologous FPP synthase.

19. The method of claim 8, wherein the host cell comprises a heterologous geranylgeranyl pyrophosphate (GGPP) synthase.

International Patent Application No. PCT/US2012/023446
Appendix B: AMENDED CLAIMS (clean version)

20. The method of claim 19, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are diterpene synthases, and the polyprenyl diphosphate is geranylgeranyl pyrophosphate (GGPP).

21. The method of claim 3, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are each selected from the group consisting of β–farnesene synthase, α–farnesene synthase, trichodiene synthase, patchoulol synthase, amorphadiene synthase, valencene synthase, farnesol synthase, nerolidol synthase, and nootkatone synthase.

22. The method of claim 1, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are each a sesterterpene synthase, a triterpene synthase, a tetraterpene synthase, or a polyterpene synthase.

23. The method of claim 1, wherein the first terpene or the second terpene is selected from the group consisting of β–farnesene, α–farnesene, trichodiene, patchoulol, amorphadiene, valencene, farnesol, nerolidol, and nootkatone.

24. The method of claim 1, wherein the first terpene or the second terpene is limonene or myrcene.

25. The method of claim 1, the method further comprising.

(e) comparing (i) the amount of the first terpene in the test population and the amount of the second terpene in the test population to (ii) the amount of the first terpene in the control population and the amount of the second terpene in the control population.

26. The method of claim 25, wherein the comparing (e) compares a first ratio to a second ratio, wherein

the first ratio is a ratio of (i) the amount of the first terpene in the test population and (ii) the amount of the second terpene in the test population, and

the second ratio is a ratio of (i) the amount of the first terpene in the control population and (ii) the amount of the second terpene in the control population.

27. The method of claim 26, wherein

the first ratio is a ratio of (i) the amount of the first terpene in the test population over (ii) the amount of the second terpene in the test population, and

the second ratio is a ratio of (i) the amount of the first terpene in the control population over (ii) the amount of the second terpene in the control population.

28. The composition of claim 14, wherein the control terpene synthase, comparison terpene synthase, and the test terpene synthase are sesquiterpene synthases, and the polyprenyl diphosphate is farnesyl diphosphate (FPP).

International Patent Application No. PCT/US2012/023446
Appendix B: AMENDED CLAIMS (clean version)

29. The composition of claim 14, wherein the control and test terpene synthases are synthases having farnesene synthase activity.

30. The composition of claim 14, wherein the comparison terpene synthase is a trichodiene synthase.

31. The composition of claim 14, wherein the host cell comprises one or more heterologous enzymes of the MEV pathway.

32. The composition of claim 14, wherein the one or more heterologous enzymes of the MEV pathway are selected from the group consisting of ERG8, ERG10, ERG12, ERG13, ERG19, and HMG1.

33. The composition of claim 14, wherein the host cell comprises a heterologous FPP synthase.