MICROORGANISMS AND METHODS FOR PRODUCING ALKENES

BACKGROUND OF THE INVENTION

[0001] This application claims the benefit of priority of United States Provisional application serial No. 61/535,893, filed September 16, 2011, the entire contents of which is incorporated herein by reference.

[0002] The present invention relates generally to biosynthetic processes, and more specifically to organisms having an alkene biosynthetic capability.

[0003] Alkenes are commonly produced by cracking the alkanes found in crude oil. Cracking uses heat and a catalyst to decompose alkanes. Generally, alkenes are unsaturated hydrocarbons with one double bond (R-C=C-R). Because of to the inherent property of alkenes being more reactive than alkanes due to the presence of a double bond, alkenes are frequently used in the manufacture of plastics. For example, alkenes are used in the maufacutre of polyethene, polyvinylchloride (PVC) and Teflon. Lower alkenes, which are obtained by the cracking of kerosene or petrol, are also commonly used as fuel and illuminant. Some alkenes, such as 1,3-butadiene, styrene and propylene, are particularly useful in manufacturing.

[0004] Over 25 billion pounds of butadiene (1,3-butadiene, BD) are produced annually and is applied in the manufacture of polymers such as synthetic rubbers and ABS resins, and chemicals such as hexamethylenediamine and 1,4-butanediol. Butadiene is typically produced as a by-product of the steam cracking process for conversion of petroleum feedstocks such as naphtha, liquefied petroleum gas, ethane or natural gas to ethylene and other olefins. The ability to manufacture butadiene from alternative and/or renewable feedstocks would represent a major advance in the quest for more sustainable chemical production processes

[0005] One possible way to produce butadiene renewably involves fermentation of sugars or other feedstocks to produce diols, such as 1,4-butanediol or 1,3-butanediol, which are separated, purified, and then dehydrated to butadiene in a second step involving metal-based catalysis. Direct fermentative production of butadiene from renewable feedstocks would obviate the need for dehydration steps and butadiene gas (bp -4.4°C) would be continuously emitted from the fermenter and readily condensed and collected.

Developing a fermentative production process would eliminate the need for fossil-based butadiene and would allow substantial savings in cost, energy, and harmful waste and emissions relative to petrochemically-derived butadiene.

[0006] Styrene is the precursor to polystyrene and numerous copolymers. Styrene based products include, acrylonitrile 1,3-butadiene styrene (ABS), styrene-1,3-butadiene (SBR) rubber, styrene-1,3-butadiene latex, SIS (styrene-isoprene-styrene), S-EB-S (styrene-ethylene/butylene-styrene), styrene-divinylbenzene (S-DVB), and unsaturated polyesters. These materials are used in rubber, plastic, insulation, fiberglass, pipes, automobile and boat parts, food containers, and carpet backing.

[0007] Styrene is most commonly produced by the catalytic dehydrogenation of ethylbenzene. Ethylbenzene is mixed in the gas phase with 10–15 times its volume in high-temperature steam, and passed over a solid catalyst bed. Most ethylbenzene dehydrogenation catalysts are based on iron(III) oxide, promoted by several percent potassium oxide or potassium carbonate. Steam serves several roles in this reaction. It is the source of heat for powering the endothermic reaction, and it removes coke that tends to form on the iron oxide catalyst through the water gas shift reaction. The potassium promoter enhances this decoking reaction. The steam also dilutes the reactant and products, shifting the position of chemical equilibrium towards products. A typical styrene plant consists of two or three reactors in series, which operate under vacuum to enhance the conversion and selectivity. Typical per-pass conversions are ca. 65% for two reactors and 70-75% for three reactors.

[0008] Propylene is produced primarily as a by-product of petroleum refining and of ethylene production by steam cracking of hydrocarbon feedstocks. Propene is separated by fractional distillation from hydrocarbon mixtures obtained from cracking and other refining processes. Typical hydrocarbon feedstocks are from non-renewable fossil fuels, such as petroleum, natural gas and to a much lesser extent coal. Over 75 billion pounds of propylene are manufactured annually, making it the second largest fossil-based chemical produced behind ethylene. Propylene is a base chemical that is converted into a wide range of polymers, polymer intermediates and chemicals. Some of the most common derivatives of chemical and polymer grade propylene are polypropylene, acrylic acid, butanol, butanediol, acrylonitrile, propylene oxide, isopropanol and cumene. The use of the propylene derivative, polypropylene, in the production of plastics, such as injection moulding, and fibers, such as carpets, accounts for over one-third of U.S. consumption for

this derivative. Propylene is also used in the production of synthetic rubber and as a propellant or component in aerosols.

[0009] The ability to manufacture propylene from alternative and/or renewable feedstocks would represent a major advance in the quest for more sustainable chemical production processes. One possible way to produce propylene renewably involves fermentation of sugars or other feedstocks to produce the alcohols 2-propanol (isopropanol) or 1-propanol, which is separated, purified, and then dehydrated to propylene in a second step involving metal-based catalysis. Direct fermentative production of propylene from renewable feedstocks would obviate the need for dehydration. During fermentative production, propylene gas would be continuously emitted from the fermenter, which could be readily collected and condensed. Developing a fermentative production process would also eliminate the need for fossil-based propylene and would allow substantial savings in cost, energy, and harmful waste and emissions relative to petrochemically-derived propylene.

[0010] Thus, there exists a need for alternative methods for effectively producing commercial quantities of alkenes. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF INVENTION

[0011] The invention provides non-naturally occurring microbial organisms containing an alkene pathway having at least one exogenous nucleic acid encoding an alkene pathway enzyme expressed in a sufficient amount to convert an alcohol to an alkene. In some aspects of the invention, the microbial organism comprises an alkene pathway selected from: (1) an alcohol kinase and a phosphate lyase; (2) a diphosphokinase and a diphosphate lyase; and (3) an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase. The invention additionally provides methods of using such microbial organisms to produce an alkene, by culturing a non-naturally occurring microbial organism containing an alkene pathway as described herein under conditions and for a sufficient period of time to produce an alkene.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 shows the conversion of an alcohol substrate to an alkene via alkyl phosphate or alkyl diphosphate intermediates. Enzymes are A. alcohol kinase, B. phosphate lyase, C. diphosphokinase, D. alkyl phosphate kinase and E. diphosphate lyase.

R^1, R^2, R^3, and R^4 are each independently (a) hydrogen, cyano, halo, or nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more substituents Q; or (c) $-C(O)R^{1a}$, $-C(O)OR^{1a}$, $-C(O)NR^{1b}R^{1c}$, $-C(NR^{1a})NR^{1b}R^{1c}$, $-OR^{1a}$, $-OC(O)R^{1a}$, $-OC(O)OR^{1a}$, $-OC(O)NR^{1b}R^{1c}$, $-OC(=NR^{1a})NR^{1b}R^{1c}$, $-OS(O)R^{1a}$, $-OS(O)_2R^{1a}$, $-OS(O)NR^{1b}R^{1c}$, $-OS(O)_2NR^{1b}R^{1c}$, $-NR^{1b}R^{1c}$, $-NR^{1a}C(O)R^{1d}$, $-NR^{1a}C(O)OR^{1d}$, $-NR^{1a}C(O)NR^{1b}R^{1c}$, $-NR^{1a}C(=NR^{1d})NR^{1b}R^{1c}$, $-NR^{1a}S(O)R^{1d}$, $-NR^{1a}S(O)_2R^{1d}$, $-NR^{1a}S(O)NR^{1b}R^{1c}$, $-NR^{1a}S(O)_2NR^{1b}R^{1c}$, $-SR^{1a}$, $-S(O)R^{1a}$, $-S(O)_2R^{1a}$, $-S(O)NR^{1b}R^{1c}$, or $-S(O)_2NR^{1b}R^{1c}$; wherein each R^{1a}, R^{1b}, R^{1c}, and R^{1d} is independently hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or R^{1a} and R^{1c} together with the C and N atoms to which they are attached form heterocyclyl; or R^{1b} and R^{1c} together with the N atom to which they are attached form heterocyclyl; wherein each Q is independently selected from (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl, each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (c) $-C(O)R^a$, $-C(O)OR^a$, $-C(O)NR^bR^c$, $-C(NR^a)NR^bR^c$, $-OR^a$, $-OC(O)R^a$, $-OC(O)OR^a$, $-OC(O)NR^bR^c$, $-OC(=NR^a)NR^bR^c$, $-OS(O)R^a$, $-OS(O)_2R^a$, $-OS(O)NR^bR^c$, $-OS(O)_2NR^bR^c$, $-NR^bR^c$, $-NR^aC(O)R^d$, $-NR^aC(O)OR^d$, $-NR^aC(O)NR^bR^c$, $-NR^aC(=NR^d)NR^bR^c$, $-NR^aS(O)R^d$, $-NR^aS(O)_2R^d$, $-NR^aS(O)NR^bR^c$, $-NR^aS(O)_2NR^bR^c$, $-SR^a$, $-S(O)R^a$, $-S(O)_2R^a$, $-S(O)NR^bR^c$, and $-S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heterocyclyl, optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a;wherein each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and

heterocyclyl; and (c) $-C(O)R^e$, $-C(O)OR^e$, $-C(O)NR^fR^g$, $-C(NR^e)NR^fR^g$, $-OR^e$, $-OC(O)R^e$, $-OC(O)OR^e$, $-OC(O)NR^fR^g$, $-OC(=NR^e)NR^fR^g$, $-OS(O)R^e$, $-OS(O)_2R^e$, $-OS(O)NR^fR^g$, $-OS(O)_2NR^fR^g$, $-NR^fR^g$, $-NR^eC(O)R^h$, $-NR^eC(O)OR^f$, $-NR^eC(O)NR^fR^g$, $-NR^eC(=NR^h)NR^fR^g$, $-NR^eS(O)R^h$, $-NR^eS(O)_2R^h$, $-NR^eS(O)NR^fR^g$, $-NR^eS(O)_2NR^fR^g$, $-SR^e$, $-S(O)R^e$, $-S(O)_2R^e$, $-S(O)NR^fR^g$, and $-S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heterocyclyl.

DETAILED DESCRIPTION OF THE INVENTION

[0013] The present invention is directed to the design and production of cells and organisms having biosynthetic production capabilities for alkenes. The invention, in particular, relates to the design of microbial organism capable of producing alkene by introducing one or more nucleic acids encoding an alkene pathway enzyme.

[0014] In one embodiment, the invention utilizes *in silico* stoichiometric models of *Escherichia coli* metabolism that identify metabolic designs for biosynthetic production of alkenes. The results described herein indicate that metabolic pathways can be designed and recombinantly engineered to achieve the biosynthesis of alkenes in *Escherichia coli* and other cells or organisms. Biosynthetic production of alkenes, for example, by the *in silico* designs can be confirmed by construction of strains having the designed metabolic genotype. These metabolically engineered cells or organisms also can be subjected to adaptive evolution to further augment alkene biosynthesis, including under conditions approaching theoretical maximum growth.

[0015] In certain embodiments, the alkene biosynthesis characteristics of the designed strains make them genetically stable and particularly useful in continuous bioprocesses. Separate strain design strategies were identified with incorporation of different non-native or heterologous reaction capabilities into *E. coli* or other host organisms leading to alkene producing metabolic pathways from alcohols that are produced naturally or that are produced through genetic engineering. *In silico* metabolic designs were identified that resulted in the biosynthesis of alkenes in microorganisms from this substrate or metabolic intermediates.

[0016] Strains identified via the computational component of the platform can be put into actual production by genetically engineering any of the predicted metabolic alterations, which lead to the biosynthetic production of alkenes or other intermediate and/or downstream products. In yet a further embodiment, strains exhibiting biosynthetic production of these compounds can be further subjected to adaptive evolution to further augment product biosynthesis. The levels of product biosynthesis yield following adaptive evolution also can be predicted by the computational component of the system.

[0017] As used herein, the term "non-naturally occurring" when used in reference to a microbial organism or microorganism of the invention is intended to mean that the microbial organism has at least one genetic alteration not normally found in a naturally occurring strain of the referenced species, including wild-type strains of the referenced species. Genetic alterations include, for example, modifications introducing expressible nucleic acids encoding metabolic polypeptides, other nucleic acid additions, nucleic acid deletions and/or other functional disruption of the microbial organism's genetic material. Such modifications include, for example, coding regions and functional fragments thereof, for heterologous, homologous or both heterologous and homologous polypeptides for the referenced species. Additional modifications include, for example, non-coding regulatory regions in which the modifications alter expression of a gene or operon. Exemplary metabolic polypeptides include enzymes or proteins within an alkene biosynthetic pathway.

[0018] A metabolic modification refers to a biochemical reaction that is altered from its naturally occurring state. Therefore, non-naturally occurring microorganisms can have genetic modifications to nucleic acids encoding metabolic polypeptides, or functional fragments thereof. Exemplary metabolic modifications are disclosed herein.

[0019] As used herein, the term "isolated" when used in reference to a microbial organism is intended to mean an organism that is substantially free of at least one component as the referenced microbial organism is found in nature. The term includes a microbial organism that is removed from some or all components as it is found in its natural environment. The term also includes a microbial organism that is removed from some or all components as the microbial organism is found in non-naturally occurring environments. Therefore, an isolated microbial organism is partly or completely separated from other substances as it is found in nature or as it is grown, stored or subsisted in non-naturally occurring environments. Specific examples of isolated microbial organisms

include partially pure microbes, substantially pure microbes and microbes cultured in a medium that is non-naturally occurring.

[0020] As used herein, the terms "microbial," "microbial organism" or "microorganism" are intended to mean any organism that exists as a microscopic cell that is included within the domains of archaea, bacteria or eukarya. Therefore, the term is intended to encompass prokaryotic or eukaryotic cells or organisms having a microscopic size and includes bacteria, archaea and eubacteria of all species as well as eukaryotic microorganisms such as yeast and fungi. The term also includes cell cultures of any species that can be cultured for the production of a biochemical.

[0021] As used herein, the term "substantially anaerobic" when used in reference to a culture or growth condition is intended to mean that the amount of oxygen is less than about 10% of saturation for dissolved oxygen in liquid media. The term also is intended to include sealed chambers of liquid or solid medium maintained with an atmosphere of less than about 1% oxygen.

[0022] "Exogenous" as it is used herein is intended to mean that the referenced molecule or the referenced activity is introduced into the host microbial organism. The molecule can be introduced, for example, by introduction of an encoding nucleic acid into the host genetic material such as by integration into a host chromosome or as non-chromosomal genetic material such as a plasmid. Therefore, the term as it is used in reference to expression of an encoding nucleic acid refers to introduction of the encoding nucleic acid in an expressible form into the microbial organism. When used in reference to a biosynthetic activity, the term refers to an activity that is introduced into the host reference organism. The source can be, for example, a homologous or heterologous encoding nucleic acid that expresses the referenced activity following introduction into the host microbial organism. Therefore, the term "endogenous" refers to a referenced molecule or activity that is present in the host. Similarly, the term when used in reference to expression of an encoding nucleic acid refers to expression of an encoding nucleic acid contained within the microbial organism. The term "heterologous" refers to a molecule or activity derived from a source other than the referenced species whereas "homologous" refers to a molecule or activity derived from the host microbial organism. Accordingly, exogenous expression of an encoding nucleic acid of the invention can utilize either or both a heterologous or homologous encoding nucleic acid.

[0023] It is understood that when more than one exogenous nucleic acid is included in a microbial organism that the more than one exogenous nucleic acids refers to the referenced encoding nucleic acid or biosynthetic activity, as discussed above. It is further understood, as disclosed herein, that such more than one exogenous nucleic acids can be introduced into the host microbial organism on separate nucleic acid molecules, on polycistronic nucleic acid molecules, or a combination thereof, and still be considered as more than one exogenous nucleic acid. For example, as disclosed herein a microbial organism can be engineered to express two or more exogenous nucleic acids encoding a desired pathway enzyme or protein. In the case where two exogenous nucleic acids encoding a desired activity are introduced into a host microbial organism, it is understood that the two exogenous nucleic acids can be introduced as a single nucleic acid, for example, on a single plasmid, on separate plasmids, can be integrated into the host chromosome at a single site or multiple sites, and still be considered as two exogenous nucleic acids. Similarly, it is understood that more than two exogenous nucleic acids can be introduced into a host organism in any desired combination, for example, on a single plasmid, on separate plasmids, can be integrated into the host chromosome at a single site or multiple sites, and still be considered as two or more exogenous nucleic acids, for example three exogenous nucleic acids. Thus, the number of referenced exogenous nucleic acids or biosynthetic activities refers to the number of encoding nucleic acids or the number of biosynthetic activities, not the number of separate nucleic acids introduced into the host organism.

[0024] The non-naturally occurring microbal organisms of the invention can contain stable genetic alterations, which refers to microorganisms that can be cultured for greater than five generations without loss of the alteration. Generally, stable genetic alterations include modifications that persist greater than 10 generations, particularly stable modifications will persist more than about 25 generations, and more particularly, stable genetic modifications will be greater than 50 generations, including indefinitely.

[0025] Those skilled in the art will understand that the genetic alterations, including metabolic modifications exemplified herein, are described with reference to a suitable host organism such as *E. coli* and their corresponding metabolic reactions or a suitable source organism for desired genetic material such as genes for a desired metabolic pathway. However, given the complete genome sequencing of a wide variety of organisms and the high level of skill in the area of genomics, those skilled in the art will readily be able to

apply the teachings and guidance provided herein to essentially all other organisms. For example, the *E. coli* metabolic alterations exemplified herein can readily be applied to other species by incorporating the same or analogous encoding nucleic acid from species other than the referenced species. Such genetic alterations include, for example, genetic alterations of species homologs, in general, and in particular, orthologs, paralogs or nonorthologous gene displacements.

[0026] An ortholog is a gene or genes that are related by vertical descent and are responsible for substantially the same or identical functions in different organisms. For example, mouse epoxide hydrolase and human epoxide hydrolase can be considered orthologs for the biological function of hydrolysis of epoxides. Genes are related by vertical descent when, for example, they share sequence similarity of sufficient amount to indicate they are homologous, or related by evolution from a common ancestor. Genes can also be considered orthologs if they share three-dimensional structure but not necessarily sequence similarity, of a sufficient amount to indicate that they have evolved from a common ancestor to the extent that the primary sequence similarity is not identifiable. Genes that are orthologous can encode proteins with sequence similarity of about 25% to 100% amino acid sequence identity. Genes encoding proteins sharing an amino acid similarity less that 25% can also be considered to have arisen by vertical descent if their three-dimensional structure also shows similarities. Members of the serine protease family of enzymes, including tissue plasminogen activator and elastase, are considered to have arisen by vertical descent from a common ancestor.

[0027] Orthologs include genes or their encoded gene products that through, for example, evolution, have diverged in structure or overall activity. For example, where one species encodes a gene product exhibiting two functions and where such functions have been separated into distinct genes in a second species, the three genes and their corresponding products are considered to be orthologs. For the production of a biochemical product, those skilled in the art will understand that the orthologous gene harboring the metabolic activity to be introduced or disrupted is to be chosen for construction of the non-naturally occurring microorganism. An example of orthologs exhibiting separable activities is where distinct activities have been separated into distinct gene products between two or more species or within a single species. A specific example is the separation of elastase proteolysis and plasminogen proteolysis, two types of serine protease activity, into distinct molecules as plasminogen activator and elastase. A second

example is the separation of mycoplasma 5'-3' exonuclease and *Drosophila* DNA polymerase III activity. The DNA polymerase from the first species can be considered an ortholog to either or both of the exonuclease or the polymerase from the second species and vice versa.

[0028] In contrast, paralogs are homologs related by, for example, duplication followed by evolutionary divergence and have similar or common, but not identical functions. Paralogs can originate or derive from, for example, the same species or from a different species. For example, microsomal epoxide hydrolase (epoxide hydrolase I) and soluble epoxide hydrolase (epoxide hydrolase II) can be considered paralogs because they represent two distinct enzymes, co-evolved from a common ancestor, that catalyze distinct reactions and have distinct functions in the same species. Paralogs are proteins from the same species with significant sequence similarity to each other suggesting that they are homologous, or related through co-evolution from a common ancestor. Groups of paralogous protein families include HipA homologs, luciferase genes, peptidases, and others.

[0029] A nonorthologous gene displacement is a nonorthologous gene from one species that can substitute for a referenced gene function in a different species. Substitution includes, for example, being able to perform substantially the same or a similar function in the species of origin compared to the referenced function in the different species. Although generally, a nonorthologous gene displacement will be identifiable as structurally related to a known gene encoding the referenced function, less structurally related but functionally similar genes and their corresponding gene products nevertheless will still fall within the meaning of the term as it is used herein. Functional similarity requires, for example, at least some structural similarity in the active site or binding region of a nonorthologous gene product compared to a gene encoding the function sought to be substituted. Therefore, a nonorthologous gene includes, for example, a paralog or an unrelated gene.

[0030] Therefore, in identifying and constructing the non-naturally occurring microbial organisms of the invention having alkene biosynthetic capability, those skilled in the art will understand with applying the teaching and guidance provided herein to a particular species that the identification of metabolic modifications can include identification and inclusion or inactivation of orthologs. To the extent that paralogs and/or nonorthologous gene displacements are present in the referenced microorganism that

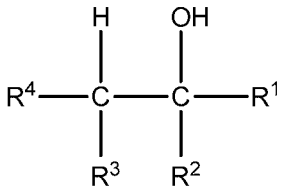
encode an enzyme catalyzing a similar or substantially similar metabolic reaction, those skilled in the art also can utilize these evolutionally related genes.

[0031] Orthologs, paralogs and nonorthologous gene displacements can be determined by methods well known to those skilled in the art. For example, inspection of nucleic acid or amino acid sequences for two polypeptides will reveal sequence identity and similarities between the compared sequences. Based on such similarities, one skilled in the art can determine if the similarity is sufficiently high to indicate the proteins are related through evolution from a common ancestor. Algorithms well known to those skilled in the art, such as Align, BLAST, Clustal W and others compare and determine a raw sequence similarity or identity, and also determine the presence or significance of gaps in the sequence which can be assigned a weight or score. Such algorithms also are known in the art and are similarly applicable for determining nucleotide sequence similarity or identity. Parameters for sufficient similarity to determine relatedness are computed based on well known methods for calculating statistical similarity, or the chance of finding a similar match in a random polypeptide, and the significance of the match determined. A computer comparison of two or more sequences can, if desired, also be optimized visually by those skilled in the art. Related gene products or proteins can be expected to have a high similarity, for example, 25% to 100% sequence identity. Proteins that are unrelated can have an identity which is essentially the same as would be expected to occur by chance, if a database of sufficient size is scanned (about 5%). Sequences between 5% and 24% may or may not represent sufficient homology to conclude that the compared sequences are related. Additional statistical analysis to determine the significance of such matches given the size of the data set can be carried out to determine the relevance of these sequences.

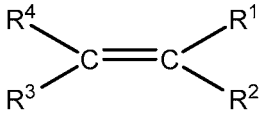
[0032] Exemplary parameters for determining relatedness of two or more sequences using the BLAST algorithm, for example, can be as set forth below. Briefly, amino acid sequence alignments can be performed using BLASTP version 2.0.8 (Jan-05-1999) and the following parameters: Matrix: 0 BLOSUM62; gap open: 11; gap extension: 1; x_dropoff: 50; expect: 10.0; wordsize: 3; filter: on. Nucleic acid sequence alignments can be performed using BLASTN version 2.0.6 (Sept-16-1998) and the following parameters: Match: 1; mismatch: -2; gap open: 5; gap extension: 2; x_dropoff: 50; expect: 10.0; wordsize: 11; filter: off. Those skilled in the art will know what modifications can be

made to the above parameters to either increase or decrease the stringency of the comparison, for example, and determine the relatedness of two or more sequences.

[0033] In some embodiments, the invention provides non-naturally occurring microbial organisms containing an alkene pathway having at least one exogenous nucleic acid encoding an alkene pathway enzyme expressed in a sufficient amount to convert an alcohol to an alkene as depicted in Figure 1. In some aspects of the invention, the microbial organism comprises an alkene pathway selected from: (1) an alcohol kinase and a phosphate lyase; (2) a diphosphokinase and a diphosphate lyase; and (3) an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase. In some aspects of the invention, the microbial organism converts an alcohol of Formula (I)



to an alkene of Formula (II)



wherein R^1, R^2, R^3, and R^4 are each independently (a) hydrogen, cyano, halo, or nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more substituents Q; or (c) $-C(O)R^{1a}$, $-C(O)OR^{1a}$, $-C(O)NR^{1b}R^{1c}$, $-C(NR^{1a})NR^{1b}R^{1c}$, $-OR^{1a}$, $-OC(O)R^{1a}$, $-OC(O)OR^{1a}$, $-OC(O)NR^{1b}R^{1c}$, $-OC(=NR^{1a})NR^{1b}R^{1c}$, $-OS(O)R^{1a}$, $-OS(O)_2R^{1a}$, $-OS(O)NR^{1b}R^{1c}$, $-OS(O)_2NR^{1b}R^{1c}$, $-NR^{1b}R^{1c}$, $-NR^{1a}C(O)R^{1d}$, $-NR^{1a}C(O)OR^{1d}$, $-NR^{1a}C(O)NR^{1b}R^{1c}$, $-NR^{1a}C(=NR^{1d})NR^{1b}R^{1c}$, $-NR^{1a}S(O)R^{1d}$, $-NR^{1a}S(O)_2R^{1d}$, $-NR^{1a}S(O)NR^{1b}R^{1c}$, $-NR^{1a}S(O)_2NR^{1b}R^{1c}$, $-SR^{1a}$, $-S(O)R^{1a}$, $-S(O)_2R^{1a}$, $-S(O)NR^{1b}R^{1c}$, or $-S(O)_2NR^{1b}R^{1c}$; wherein each R^{1a}, R^{1b}, R^{1c}, and R^{1d} is independently hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or R^{1a} and R^{1c} together with the C and N atoms to which they are attached form heterocyclyl; or R^{1b} and R^{1c} together with the N atom to which they are attached form heterocyclyl; wherein each Q is independently selected from (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl,

each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (c) $-C(O)R^a$, $-C(O)OR^a$, $-C(O)NR^bR^c$, $-C(NR^a)NR^bR^c$, $-OR^a$, $-OC(O)R^a$, $-OC(O)OR^a$, $-OC(O)NR^bR^c$, $-OC(=NR^a)NR^bR^c$, $-OS(O)R^a$, $-OS(O)_2R^a$, $-OS(O)NR^bR^c$, $-OS(O)_2NR^bR^c$, $-NR^bR^c$, $-NR^aC(O)R^d$, $-NR^aC(O)OR^d$, $-NR^aC(O)NR^bR^c$, $-NR^aC(=NR^d)NR^bR^c$, $-NR^aS(O)R^d$, $-NR^aS(O)_2R^d$, $-NR^aS(O)NR^bR^c$, $-NR^aS(O)_2NR^bR^c$, $-SR^a$, $-S(O)R^a$, $-S(O)_2R^a$, $-S(O)NR^bR^c$, and $-S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heterocyclyl, optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; wherein each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl; and (c) $-C(O)R^e$, $-C(O)OR^e$, $-C(O)NR^fR^g$, $-C(NR^e)NR^fR^g$, $-OR^e$, $-OC(O)R^e$, $-OC(O)OR^e$, $-OC(O)NR^fR^g$, $-OC(=NR^e)NR^fR^g$, $-OS(O)R^e$, $-OS(O)_2R^e$, $-OS(O)NR^fR^g$, $-OS(O)_2NR^fR^g$, $-NR^fR^g$, $-NR^eC(O)R^h$, $-NR^eC(O)OR^f$, $-NR^eC(O)NR^fR^g$, $-NR^eC(=NR^h)NR^fR^g$, $-NR^eS(O)R^h$, $-NR^eS(O)_2R^h$, $-NR^eS(O)NR^fR^g$, $-NR^eS(O)_2NR^fR^g$, $-SR^e$, $-S(O)R^e$, $-S(O)_2R^e$, $-S(O)NR^fR^g$, and $-S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heterocyclyl. It is also understood that R^1, R^2, R^3, and R^4 are each independently same between the alcohol and the alkene. In otherwords, the R^1 of the alcohol is the same as the R^1 of the alkene, the R^2 of the alcohol is the same as the R^2 of the alkene, the R^3 of the alcohol is the same as the R^3 of the alkene and the R^4 of the alcohol is the same as the R^4 of the alkene.

[0034] The term "alkyl" refers to a linear or branched saturated monovalent hydrocarbon radical, wherein the alkyl may optionally be substituted with one or more substituents Q as described herein. For example, C_{1-6} alkyl refers to a linear saturated monovalent hydrocarbon radical of 1 to 6 carbon atoms or a branched saturated monovalent hydrocarbon radical of 3 to 6 carbon atoms. In certain embodiments, the alkyl is a linear saturated monovalent hydrocarbon radical that has 1 to 20 (C_{1-20}), 1 to 15 (C_{1-15}), 1 to 10 (C_{1-10}), or 1 to 6 (C_{1-6}) carbon atoms, or branched saturated monovalent

hydrocarbon radical of 3 to 20 (C_{3-20}), 3 to 15 (C_{3-15}), 3 to 10 (C_{3-10}), or 3 to 6 (C_{3-6}) carbon atoms. As used herein, linear C_{1-6} and branched C_{3-6} alkyl groups are also referred as "lower alkyl." Examples of alkyl groups include, but are not limited to, methyl, ethyl, propyl (including all isomeric forms), *n*-propyl, isopropyl, butyl (including all isomeric forms), *n*-butyl, isobutyl, *sec*-butyl, *t*-butyl, pentyl (including all isomeric forms), and hexyl (including all isomeric forms).

[0035] The term "alkenyl" refers to a linear or branched monovalent hydrocarbon radical, which contains one or more, in one embodiment, one to five, in another embodiment, one, carbon-carbon double bond(s). The alkenyl may be optionally substituted with one or more substituents Q as described herein. The term "alkenyl" embraces radicals having a "*cis*" or "*trans*" configuration or a mixture thereof, or alternatively, a "Z" or "E" configuration or a mixture thereof, as appreciated by those of ordinary skill in the art. For example, C_{2-6} alkenyl refers to a linear unsaturated monovalent hydrocarbon radical of 2 to 6 carbon atoms or a branched unsaturated monovalent hydrocarbon radical of 3 to 6 carbon atoms. In certain embodiments, the alkenyl is a linear monovalent hydrocarbon radical of 2 to 20 (C_{2-20}), 2 to 15 (C_{2-15}), 2 to 10 (C_{2-10}), or 2 to 6 (C_{2-6}) carbon atoms, or a branched monovalent hydrocarbon radical of 3 to 20 (C_{3-20}), 3 to 15 (C_{3-15}), 3 to 10 (C_{3-10}), or 3 to 6 (C_{3-6}) carbon atoms. Examples of alkenyl groups include, but are not limited to, ethenyl, propen-1-yl, propen-2-yl, allyl, butenyl, and 4-methylbutenyl.

[0036] The term "alkynyl" refers to a linear or branched monovalent hydrocarbon radical, which contains one or more, in one embodiment, one to five, in another embodiment, one, carbon-carbon triple bond(s). The alkynyl may be optionally substituted with one or more substituents Q as described herein. For example, C_{2-6} alkynyl refers to a linear unsaturated monovalent hydrocarbon radical of 2 to 6 carbon atoms or a branched unsaturated monovalent hydrocarbon radical of 3 to 6 carbon atoms. In certain embodiments, the alkynyl is a linear monovalent hydrocarbon radical of 2 to 20 (C_{2-20}), 2 to 15 (C_{2-15}), 2 to 10 (C_{2-10}), or 2 to 6 (C_{2-6}) carbon atoms, or a branched monovalent hydrocarbon radical of 3 to 20 (C_{3-20}), 3 to 15 (C_{3-15}), 3 to 10 (C_{3-10}), or 3 to 6 (C_{3-6}) carbon atoms. Examples of alkynyl groups include, but are not limited to, ethynyl (–C≡CH), propynyl (including all isomeric forms, *e.g.*, 1-propynyl (–C≡CCH_3) and propargyl (–CH_2C≡CH)), butynyl (including all isomeric forms, *e.g.*, 1-butyn-1-yl and 2-butyn-1-yl),

pentynyl (including all isomeric forms, *e.g.*, 1-pentyn-1-yl and 1-methyl-2-butyn-1-yl), and hexynyl (including all isomeric forms, *e.g.*, 1-hexyn-1-yl).

[0037] The term "cycloalkyl" refers to a cyclic monovalent hydrocarbon radical, which may be optionally substituted with one or more substituents Q as described herein. In one embodiment, cycloalkyl groups may be saturated or unsaturated but non-aromatic, and/or bridged, and/or non-bridged, and/or fused bicyclic groups. In certain embodiments, the cycloalkyl has from 3 to 20 (C_{3-20}), from 3 to 15 (C_{3-15}), from 3 to 10 (C_{3-10}), or from 3 to 7 (C_{3-7}) carbon atoms. Examples of cycloalkyl groups include, but are not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclopentenyl, cyclohexyl, cyclohexenyl, cyclohexadienyl, cycloheptyl, cycloheptenyl, bicyclo[2.1.1]hexyl, bicyclo[2.2.1]heptyl, decalinyl, and adamantyl.

[0038] The term "aryl" refers to a monovalent monocyclic aromatic group and/or monovalent polycyclic aromatic group that contain at least one aromatic carbon ring. In certain embodiments, the aryl has from 6 to 20 (C_{6-20}), from 6 to 15 (C_{6-15}), or from 6 to 10 (C_{6-10}) ring atoms. Examples of aryl groups include, but are not limited to, phenyl, naphthyl, fluorenyl, azulenyl, anthryl, phenanthryl, pyrenyl, biphenyl, and terphenyl. Aryl also refers to bicyclic or tricyclic carbon rings, where one of the rings is aromatic and the others of which may be saturated, partially unsaturated, or aromatic, for example, dihydronaphthyl, indenyl, indanyl, or tetrahydronaphthyl (tetralinyl). In certain embodiments, aryl may be optionally substituted with one or more substituents Q as described herein.

[0039] The term "aralkyl" or "arylalkyl" refers to a monovalent alkyl group substituted with one or more aryl groups. In certain embodiments, the aralkyl has from 7 to 30 (C_{7-30}), from 7 to 20 (C_{7-20}), or from 7 to 16 (C_{7-16}) carbon atoms. Examples of aralkyl groups include, but are not limited to, benzyl, 2-phenylethyl, and 3-phenylpropyl. In certain embodiments, aralkyl are optionally substituted with one or more substituents Q as described herein.

[0040] The term "heteroaryl" refers to a monovalent monocyclic aromatic group or monovalent polycyclic aromatic group that contain at least one aromatic ring, wherein at least one aromatic ring contains one or more heteroatoms independently selected from O, S, and N in the ring. Heteroaryl groups are bonded to the rest of a molecule through the aromatic ring. Each ring of a heteroaryl group can contain one or two O atoms, one or two

S atoms, and/or one to four N atoms, provided that the total number of heteroatoms in each ring is four or less and each ring contains at least one carbon atom. In certain embodiments, the heteroaryl has from 5 to 20, from 5 to 15, or from 5 to 10 ring atoms. Examples of monocyclic heteroaryl groups include, but are not limited to, furanyl, imidazolyl, isothiazolyl, isoxazolyl, oxadiazolyl, oxadiazolyl, oxazolyl, pyrazinyl, pyrazolyl, pyridazinyl, pyridyl, pyrimidinyl, pyrrolyl, thiadiazolyl, thiazolyl, thienyl, tetrazolyl, triazinyl, and triazolyl. Examples of bicyclic heteroaryl groups include, but are not limited to, benzofuranyl, benzimidazolyl, benzoisoxazolyl, benzopyranyl, benzothiadiazolyl, benzothiazolyl, benzothienyl, benzotriazolyl, benzoxazolyl, furopyridyl, imidazopyridinyl, imidazothiazolyl, indolizinyl, indolyl, indazolyl, isobenzofuranyl, isobenzothienyl, isoindolyl, isoquinolinyl, isothiazolyl, naphthyridinyl, oxazolopyridinyl, phthalazinyl, pteridinyl, purinyl, pyridopyridyl, pyrrolopyridyl, quinolinyl, quinoxalinyl, quinazolinyl, thiadiazolopyrimidyl, and thienopyridyl. Examples of tricyclic heteroaryl groups include, but are not limited to, acridinyl, benzindolyl, carbazolyl, dibenzofuranyl, perimidinyl, phenanthrolinyl, phenanthridinyl, phenarsazinyl, phenazinyl, phenothiazinyl, phenoxazinyl, and xanthenyl. In certain embodiments, heteroaryl may also be optionally substituted with one or more substituents Q as described herein.

[0041] The term "heterocyclyl" or "heterocyclic" refers to a monovalent monocyclic non-aromatic ring system or monovalent polycyclic ring system that contains at least one non-aromatic ring, wherein one or more of the non-aromatic ring atoms are heteroatoms independently selected from O, S, and N; and the remaining ring atoms are carbon atoms. In certain embodiments, the heterocyclyl or heterocyclic group has from 3 to 20, from 3 to 15, from 3 to 10, from 3 to 8, from 4 to 7, or from 5 to 6 ring atoms. Heterocyclyl groups are bonded to the rest of a molecule through the non-aromatic ring. In certain embodiments, the heterocyclyl is a monocyclic, bicyclic, tricyclic, or tetracyclic ring system, which may be fused or bridged, and in which nitrogen or sulfur atoms may be optionally oxidized, nitrogen atoms may be optionally quaternized, and some rings may be partially or fully saturated, or aromatic. The heterocyclyl may be attached to the main structure at any heteroatom or carbon atom which results in the creation of a stable compound. Examples of such heterocyclic groups include, but are not limited to, azepinyl, benzodioxanyl, benzodioxolyl, benzofuranonyl, benzopyranonyl, benzopyranyl, benzotetrahydrofuranyl, benzotetrahydrothienyl, benzothiopyranyl, benzoxazinyl, β-carbolinyl, chromanyl, chromonyl, cinnolinyl, coumarinyl, decahydroisoquinolinyl,

dihydrobenzisothiazinyl, dihydrobenzisoxazinyl, dihydrofuryl, dihydroisoindolyl, dihydropyranyl, dihydropyrazolyl, dihydropyrazinyl, dihydropyridinyl, dihydropyrimidinyl, dihydropyrrolyl, dioxolanyl, 1,4-dithianyl, furanonyl, imidazolidinyl, imidazolinyl, indolinyl, isobenzotetrahydrofuranyl, isobenzotetrahydrothienyl, isochromanyl, isocoumarinyl, isoindolinyl, isothiazolidinyl, isoxazolidinyl, morpholinyl, octahydroindolyl, octahydroisoindolyl, oxazolidinonyl, oxazolidinyl, oxiranyl, piperazinyl, piperidinyl, 4-piperidonyl, pyrazolidinyl, pyrazolinyl, pyrrolidinyl, pyrrolinyl, quinuclidinyl, tetrahydrofuryl, tetrahydroisoquinolinyl, tetrahydropyranyl, tetrahydrothienyl, thiamorpholinyl, thiazolidinyl, tetrahydroquinolinyl, and 1,3,5-trithianyl. In certain embodiments, heterocyclic may also be optionally substituted with one or more substituents Q as described herein.

[0042] The term "halogen", "halide" or "halo" refers to fluorine, chlorine, bromine, and/or iodine.

[0043] The term "optionally substituted" is intended to mean that a group or substituent, such as an alkyl, alkylene, heteroalkylene, alkenyl, alkenylene, heteroalkenylene, alkynyl, alkynylene, cycloalkyl, cycloalkylene, aryl, arylene, aralkyl, heteroaryl, heteroarylene, heterocyclyl, or heterocyclylene group, may be substituted with one or more substituents Q, each of which is independently selected from, *e.g.*, (a) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl, each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (b) oxo (=O), halo, cyano (–CN), nitro ($–NO_2$), $–C(O)R^a$, $–C(O)OR^a$, $–C(O)NR^bR^c$, $–C(NR^a)NR^bR^c$, $–OR^a$, $–OC(O)R^a$, $–OC(O)OR^a$, $–OC(O)NR^bR^c$, $–OC(=NR^a)NR^bR^c$, $–OS(O)R^a$, $–OS(O)_2R^a$, $–OS(O)NR^bR^c$, $–OS(O)_2NR^bR^c$, $–NR^bR^c$, $–NR^aC(O)R^d$, $–NR^aC(O)OR^d$, $–NR^aC(O)NR^bR^c$, $–NR^aC(=NR^d)NR^bR^c$, $–NR^aS(O)R^d$, $–NR^aS(O)_2R^d$, $–NR^aS(O)NR^bR^c$, $–NR^aS(O)_2NR^bR^c$, $–SR^a$, $–S(O)R^a$, $–S(O)_2R^a$, $–S(O)NR^bR^c$, and $–S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heteroaryl or heterocyclyl, optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a. As used herein, all groups that can be substituted are "optionally substituted," unless otherwise specified.

[0044] In one embodiment, each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; and (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl; and (c) $–C(O)R^e$, $–C(O)OR^e$, $–C(O)NR^fR^g$, $–C(NR^e)NR^fR^g$, $–OR^e$, $–OC(O)R^e$, $–OC(O)OR^e$, $–OC(O)NR^fR^g$, $–OC(=NR^e)NR^fR^g$, $–OS(O)R^e$, $–OS(O)_2R^e$, $–OS(O)NR^fR^g$, $–OS(O)_2NR^fR^g$, $–NR^fR^g$, $–NR^eC(O)R^h$, $–NR^eC(O)OR^h$, $–NR^eC(O)NR^fR^g$, $–NR^eC(=NR^h)NR^fR^g$, $–NR^eS(O)R^h$, $–NR^eS(O)_2R^h$, $–NR^eS(O)NR^fR^g$, $–NR^eS(O)_2NR^fR^g$, $–SR^e$, $–S(O)R^e$, $–S(O)_2R^e$, $–S(O)NR^fR^g$, and $–S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heteroaryl or heterocyclyl.

[0045] Figure 1 shows pathways for converting an alcohol to an alkene via a phosphate or diphosphate intermediate. In step A, an alcohol is activated to an alkyl phosphate by a kinase. The alkyl phosphate is then further activated to an alkyl diphosphate (Step D) or converted to an alkene by a phosphate lyase or alkene synthase (step B). Alternately, the alcohol is directly converted to the alkyl diphosphate intermediate by a diphosphokinse (step C). The release of diphosphate from alkyl diphosphate by an alkene synthase or diphosphate lyase yields an alkene. Exemplary alcohol precursors and alkene products are listed in the table below.

Alcohol	**Alkene**
Ethanol	Ethylene
n-Propanol	Propylene
Isopropanol	Propylene
n-Butanol	But-1-ene
Isobutanol	Isobutylene
Tert-butanol	Isobutylene
Butan-2-ol	But-1-ene or but-2-ene
Pentan-1-ol	Pent-1-ene
3-methylbutan-1-ol	3-methylbut-1-ene
Pentan-2-ol	Pent-2-ene or pent-2-ene
Pentan-3-ol	Pent-2-ene
2-Methylbutan-1-ol	2-methylbut-1-ene
3-Methylbutan-2-ol	3-Methylbut-1-ene
2-Methylbutan-2-ol	2-Methylbut-1-ene or 2-Methylbut-2-ene
3-Methylbut-3-en-1-ol	Isoprene
2-Methylbut-3-en-2-ol	Isoprene

Alcohol	Alkene
2-Methylbut-3-en-2-ol	3-Methylbuta-1,2-diene
2-Methylbut-3-en-1-ol	Isoprene
3-Methylbut-3-en-2-ol	Isoprene
But-3-en-1-ol	1,3-Butadiene
But-3-en-2-ol	1,3-Butadiene
1-Phenylethanol	Styrene
2-Phenylethanol	Styrene
Dimethylallyl alcohol	Isoprene
But-2-en-1-ol	1,3-Butadiene

[0046] Accordingly, in some aspects, the invention provides a non-naturally occurring microbial organism containing an alkene pathway having at least one exogenous nucleic acid encoding an alkene pathway enzyme expressed in a sufficient amount to convert an alcohol to an alkene as depicted in the table above. In some aspects of the invention, the microbial organism comprises an alkene pathway selected from: (1) an alcohol kinase and a phosphate lyase; (2) a diphosphokinase and a diphosphate lyase; and (3) an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase. In some aspects, the microbial orgainism of the invention converts ethanol to ethylene, n-propanol to propylene, isopropanol to propylene, n-butanol to but-1-ene, isobutanol to isobutylene, tert-butanol to isobutylene, butan-2-ol to but-1-ene or but-2-ene, pentan-1-ol to pent-1-ene, 3-methylbutan-1-ol to 3-methylbut-1-ene, pentan-2-ol to pent-2-ene, pental-3-ol to pent-2-ene, 2-methylbutan-1-ol to 2-methylbut-1-ene, 3-methylbutan-2-ol to 3-methylbut-1-ene, 2-methylbutan-2-ol to 2-methylbut-1-ene or 2-methylbut-2-ene, 3-methylbut-3-en-1-ol to isoprene, 2-methylbut-3-en-2-ol to isoprene, 2-methylbut-3-en-2-ol to 3-methylbuta-1,2-diene, 2-methylbut-3-en-1-ol to isoprene, 3-methylbut-3-en-2-ol to isoprene, but-3-en-1-ol to 1,3-butadiene, but-3-en-2-ol to 1,3-butadiene, 1-phenylethanol to styrene, 2-phenylethanol to styrene, dimethylallyl alcohol to isoprene, or but-2-en-1-ol to 1,3-butadiene.

[0047] In some embodiments of the invention, the non-naturally occurring microbial organism comprises two or three exogenous nucleic acids each encoding an alkene pathway enzyme. For example, two exogenous nucleic acids can encode an alcohol kinase and a phosphate lyase, or alternatively a diphosphokinase and a diphosphate lyase. In some aspects of the invention, non-naturally occurring microbial organism can include three exogenous nucleic acids encoding an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase. The invention also provides that the at least one exogenous nucleic

acid can be a heterologous nucleic acid. The invention still further provides that the non-naturally occurring microbial organism can be in a substantially anaerobic culture medium.

[0048] In an additional embodiment, the invention provides a non-naturally occurring microbial organism having an alkene pathway, wherein the non-naturally occurring microbial organism comprises at least one exogenous nucleic acid encoding an enzyme or protein that converts a substrate to a product selected from the group consisting of an alcohol to an alkyl phosphate, an alcohol to an alkyl diphosphate, an alkyl phosphate to an alkyl diphosphate, an alkyl phosphate to an alkene or an alkyl diphosphate to an alkene. One skilled in the art will understand that these are merely exemplary and that any of the substrate-product pairs disclosed herein suitable to produce a desired product and for which an appropriate activity is available for the conversion of the substrate to the product can be readily determined by one skilled in the art based on the teachings herein. Thus, the invention provides a non-naturally occurring microbial organism containing at least one exogenous nucleic acid encoding an enzyme or protein, where the enzyme or protein converts the substrates and products of an alkene pathway, such as that shown in Figure 1.

[0049] While generally described herein as a microbial organism that contains an alkene pathway, it is understood that the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding an alkene pathway enzyme expressed in a sufficient amount to produce an intermediate of an alkene pathway. For example, as disclosed herein, an alkene pathway is exemplified in Figure 1. Therefore, in addition to a microbial organism containing an alkene pathway that produces alkene, the invention additionally provides a non-naturally occurring microbial organism comprising at least one exogenous nucleic acid encoding an alkene pathway enzyme, where the microbial organism produces an alkene pathway intermediate, for example, an alkyl phosphate or an alkyl diphosphate.

[0050] It is understood that any of the pathways disclosed herein, as described in the Examples and exemplified in the Figures, including the pathways of Figure 1, can be utilized to generate a non-naturally occurring microbial organism that produces any pathway intermediate or product, as desired. As disclosed herein, such a microbial organism that produces an intermediate can be used in combination with another microbial organism expressing downstream pathway enzymes to produce a desired product. However, it is understood that a non-naturally occurring microbial organism that produces

an alkene pathway intermediate can be utilized to produce the intermediate as a desired product.

[0051] The invention is described herein with general reference to the metabolic reaction, reactant or product thereof, or with specific reference to one or more nucleic acids or genes encoding an enzyme associated with or catalyzing, or a protein associated with, the referenced metabolic reaction, reactant or product. Unless otherwise expressly stated herein, those skilled in the art will understand that reference to a reaction also constitutes reference to the reactants and products of the reaction. Similarly, unless otherwise expressly stated herein, reference to a reactant or product also references the reaction, and reference to any of these metabolic constituents also references the gene or genes encoding the enzymes that catalyze or proteins involved in the referenced reaction, reactant or product. Likewise, given the well known fields of metabolic biochemistry, enzymology and genomics, reference herein to a gene or encoding nucleic acid also constitutes a reference to the corresponding encoded enzyme and the reaction it catalyzes or a protein associated with the reaction as well as the reactants and products of the reaction.

[0052] The non-naturally occurring microbial organisms of the invention can be produced by introducing expressible nucleic acids encoding one or more of the enzymes or proteins participating in one or more alkene biosynthetic pathways. Depending on the host microbial organism chosen for biosynthesis, nucleic acids for some or all of a particular alkene biosynthetic pathway can be expressed. For example, if a chosen host is deficient in one or more enzymes or proteins for a desired biosynthetic pathway, then expressible nucleic acids for the deficient enzyme(s) or protein(s) are introduced into the host for subsequent exogenous expression. Alternatively, if the chosen host exhibits endogenous expression of some pathway genes, but is deficient in others, then an encoding nucleic acid is needed for the deficient enzyme(s) or protein(s) to achieve alkene biosynthesis. Thus, a non-naturally occurring microbial organism of the invention can be produced by introducing exogenous enzyme or protein activities to obtain a desired biosynthetic pathway or a desired biosynthetic pathway can be obtained by introducing one or more exogenous enzyme or protein activities that, together with one or more endogenous enzymes or proteins, produces a desired product such as alkene.

[0053] Host microbial organisms can be selected from, and the non-naturally occurring microbial organisms generated in, for example, bacteria, yeast, fungus or any of

a variety of other microorganisms applicable to fermentation processes. Exemplary bacteria include species selected from *Escherichia coli, Klebsiella oxytoca, Anaerobiospirillum succiniciproducens, Actinobacillus succinogenes, Mannheimia succiniciproducens, Rhizobium etli, Bacillus subtilis, Corynebacterium glutamicum, Gluconobacter oxydans, Zymomonas mobilis, Lactococcus lactis, Lactobacillus plantarum, Streptomyces coelicolor, Clostridium acetobutylicum, Pseudomonas fluorescens*, and *Pseudomonas putida*. Exemplary yeasts or fungi include species selected from *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces lactis, Kluyveromyces marxianus, Aspergillus terreus, Aspergillus niger*, *Pichia pastoris, Rhizopus arrhizus, Rhizobus oryzae*, *Yarrowia lipolytica*, and the like. *E. coli* is a particularly useful host organism since it is a well characterized microbial organism suitable for genetic engineering. Other particularly useful host organisms include yeast such as *Saccharomyces cerevisiae*. It is understood that any suitable microbial host organism can be used to introduce metabolic and/or genetic modifications to produce a desired product.

[0054] Depending on the alkene biosynthetic pathway constituents of a selected host microbial organism, the non-naturally occurring microbial organisms of the invention will include at least one exogenously expressed alkene pathway-encoding nucleic acid and up to all encoding nucleic acids for one or more alkene biosynthetic pathways. For example, alkene biosynthesis can be established in a host deficient in a pathway enzyme or protein through exogenous expression of the corresponding encoding nucleic acid. In a host deficient in all enzymes or proteins of an alkene pathway, exogenous expression of all enzyme or proteins in the pathway can be included, although it is understood that all enzymes or proteins of a pathway can be expressed even if the host contains at least one of the pathway enzymes or proteins. For example, exogenous expression of all enzymes or proteins in a pathway for production of alkene can be included, such as an alcohol kinase and a phosphate lyase, or alternatively a diphosephokinase and a diphosphate lyase, or alternatively an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase.

[0055] Given the teachings and guidance provided herein, those skilled in the art will understand that the number of encoding nucleic acids to introduce in an expressible form will, at least, parallel the alkene pathway deficiencies of the selected host microbial organism. Therefore, a non-naturally occurring microbial organism of the invention can have one, two or three up to all nucleic acids encoding the enzymes or proteins

constituting an alkene biosynthetic pathway disclosed herein. In some embodiments, the non-naturally occurring microbial organisms also can include other genetic modifications that facilitate or optimize alkene biosynthesis or that confer other useful functions onto the host microbial organism. One such other functionality can include, for example, augmentation of the synthesis of one or more of the alkene pathway precursors such as an alcohol disclosed herein.

[0056] Generally, a host microbial organism is selected such that it produces the precursor of an alkene pathway, either as a naturally produced molecule or as an engineered product that either provides *de novo* production of a desired precursor or increased production of a precursor naturally produced by the host microbial organism. For example, ethanol is produced naturally in a host organism such as *E. coli*. A host organism can be engineered to increase production of a precursor, as disclosed herein. In addition, a microbial organism that has been engineered to produce a desired precursor can be used as a host organism and further engineered to express enzymes or proteins of an alkene pathway.

[0057] In some embodiments, a non-naturally occurring microbial organism of the invention is generated from a host that contains the enzymatic capability to synthesize alkene. In this specific embodiment it can be useful to increase the synthesis or accumulation of an alkene pathway product to, for example, drive alkene pathway reactions toward alkene production. Increased synthesis or accumulation can be accomplished by, for example, overexpression of nucleic acids encoding one or more of the above-described alkene pathway enzymes or proteins. Overexpression of the enzyme or enzymes and/or protein or proteins of the alkene pathway can occur, for example, through exogenous expression of the endogenous gene or genes, or through exogenous expression of the heterologous gene or genes. Therefore, naturally occurring organisms can be readily generated to be non-naturally occurring microbial organisms of the invention, for example, producing alkene, through overexpression of one, two, or three, that is, up to all nucleic acids encoding alkene biosynthetic pathway enzymes or proteins. In addition, a non-naturally occurring organism can be generated by mutagenesis of an endogenous gene that results in an increase in activity of an enzyme in the alkene biosynthetic pathway.

[0058] In particularly useful embodiments, exogenous expression of the encoding nucleic acids is employed. Exogenous expression confers the ability to custom tailor the

expression and/or regulatory elements to the host and application to achieve a desired expression level that is controlled by the user. However, endogenous expression also can be utilized in other embodiments such as by removing a negative regulatory effector or induction of the gene's promoter when linked to an inducible promoter or other regulatory element. Thus, an endogenous gene having a naturally occurring inducible promoter can be up-regulated by providing the appropriate inducing agent, or the regulatory region of an endogenous gene can be engineered to incorporate an inducible regulatory element, thereby allowing the regulation of increased expression of an endogenous gene at a desired time. Similarly, an inducible promoter can be included as a regulatory element for an exogenous gene introduced into a non-naturally occurring microbial organism.

[0059] It is understood that, in methods of the invention, any of the one or more exogenous nucleic acids can be introduced into a microbial organism to produce a non-naturally occurring microbial organism of the invention. The nucleic acids can be introduced so as to confer, for example, an alkene biosynthetic pathway onto the microbial organism. Alternatively, encoding nucleic acids can be introduced to produce an intermediate microbial organism having the biosynthetic capability to catalyze some of the required reactions to confer alkene biosynthetic capability. For example, a non-naturally occurring microbial organism having an alkene biosynthetic pathway can comprise at least two exogenous nucleic acids encoding desired enzymes or proteins, such as the combination of an alcohol kinase and a phosphate lyase, or alternatively a diphosphokinase and a diphosphate lyase, and the like. Thus, it is understood that any combination of two or more enzymes or proteins of a biosynthetic pathway can be included in a non-naturally occurring microbial organism of the invention. Similarly, it is understood that any combination of three or more enzymes or proteins of a biosynthetic pathway can be included in a non-naturally occurring microbial organism of the invention, for example, an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase, and so forth, as desired, so long as the combination of enzymes and/or proteins of the desired biosynthetic pathway results in production of the corresponding desired product.

[0060] In addition to the biosynthesis of alkene as described herein, the non-naturally occurring microbial organisms and methods of the invention also can be utilized in various combinations with each other and with other microbial organisms and methods well known in the art to achieve product biosynthesis by other routes. For example, one alternative to produce alkene other than use of the alkene producers is through addition of

another microbial organism capable of converting an alkene pathway intermediate to alkene. One such procedure includes, for example, the fermentation of a microbial organism that produces an alkene pathway intermediate. The alkene pathway intermediate can then be used as a substrate for a second microbial organism that converts the alkene pathway intermediate to alkene. The alkene pathway intermediate can be added directly to another culture of the second organism or the original culture of the alkene pathway intermediate producers can be depleted of these microbial organisms by, for example, cell separation, and then subsequent addition of the second organism to the fermentation broth can be utilized to produce the final product without intermediate purification steps.

[0061] In other embodiments, the non-naturally occurring microbial organisms and methods of the invention can be assembled in a wide variety of subpathways to achieve biosynthesis of, for example, alkene. In these embodiments, biosynthetic pathways for a desired product of the invention can be segregated into different microbial organisms, and the different microbial organisms can be co-cultured to produce the final product. In such a biosynthetic scheme, the product of one microbial organism is the substrate for a second microbial organism until the final product is synthesized. For example, the biosynthesis of alkene can be accomplished by constructing a microbial organism that contains biosynthetic pathways for conversion of one pathway intermediate to another pathway intermediate or the product. Alternatively, alkene also can be biosynthetically produced from microbial organisms through co-culture or co-fermentation using two organisms in the same vessel, where the first microbial organism produces an alkyl phosphate or alkyl diphosphate intermediate and the second microbial organism converts the intermediate to alkene.

[0062] Given the teachings and guidance provided herein, those skilled in the art will understand that a wide variety of combinations and permutations exist for the non-naturally occurring microbial organisms and methods of the invention together with other microbial organisms, with the co-culture of other non-naturally occurring microbial organisms having subpathways and with combinations of other chemical and/or biochemical procedures well known in the art to produce alkene.

[0063] Sources of encoding nucleic acids for an alkene pathway enzyme or protein can include, for example, any species where the encoded gene product is capable of catalyzing the referenced reaction. Such species include both prokaryotic and eukaryotic organisms including, but not limited to, bacteria, including archaea and eubacteria, and eukaryotes,

including yeast, plant, insect, animal, and mammal, including human. Exemplary species for such sources include, for example, *Escherichia coli, Abies grandis, Acetobacter pasteurians, Acinetobacter sp. strain M-1, Arabidopsis thaliana, Arabidopsis thaliana col, Aspergillus terreus NIH2624, Bacillus amyloliquefaciens , Bacillus cereus, Bos Taurus, Bradyrhizobium japonicum USDA110, Burkholderia phymatum, Burkholderia xenovorans, Clostridium acetobutylicum, Clostridium beijerinckii, Clostridium beijerinckii NRRL B593, Clostridium botulinum, Clostridium kluyveri DSM 555, Clostridium saccharoperbutylacetonicum, Comamonas sp. CNB-1, Cucumis sativus, Cupriavidus taiwanensis, Enterococcus faecalis, Escherichia coli C, Escherichia coli K12, Escherichia coli W, Geobacillus thermoglucosidasius , Homo sapiens, Klebsiella pneumonia, Kluyveromyces lactis, Lactococcus lactis, Malus x domestica, Mesorhizobium loti, Methanocaldococcus jannaschii, Methanosarcina mazei, Mycobacterium tuberculosis, Mycoplasma pneumoniae M129, Neurospora crassa, Oryctolagus cuniculus, Picea abies, Populus alba, Populus tremula x Populus alba, Pseudomonas aeruginosa, Pseudomonas putida, Pseudomonas sp. CF600, Pueraria Montana, Pyrococcus furiosus, Ralstonia eutropha, Ralstonia eutropha H16, Ralstonia metallidurans, Rattus norvegicus, Rhodococcus ruber, Saccharomyces cerevisiae, Salmonella enteric, Solanum lycopersicum, Staphylococcus aureus, Streptococcus pneumonia, Streptomyces sp. ACT-1, Thermoanaerobacter brockii HTD4, Thermotoga maritime MSB8, Thermus thermophilus, Zea mays, Zoogloea ramigera, Zymomonas mobilis*, as well as other exemplary species disclosed herein or available as source organisms for corresponding genes. However, with the complete genome sequence available for now more than 550 species (with more than half of these available on public databases such as the NCBI), including 395 microorganism genomes and a variety of yeast, fungi, plant, and mammalian genomes, the identification of genes encoding the requisite alkene biosynthetic activity for one or more genes in related or distant species, including for example, homologues, orthologs, paralogs and nonorthologous gene displacements of known genes, and the interchange of genetic alterations between organisms is routine and well known in the art. Accordingly, the metabolic alterations allowing biosynthesis of alkene described herein with reference to a particular organism such as *E. coli* can be readily applied to other microorganisms, including prokaryotic and eukaryotic organisms alike. Given the teachings and guidance provided herein, those skilled in the art will know that a metabolic alteration exemplified in one organism can be applied equally to other organisms.

[0064] In some instances, such as when an alternative alkene biosynthetic pathway exists in an unrelated species, alkene biosynthesis can be conferred onto the host species by, for example, exogenous expression of a paralog or paralogs from the unrelated species that catalyzes a similar, yet non-identical metabolic reaction to replace the referenced reaction. Because certain differences among metabolic networks exist between different organisms, those skilled in the art will understand that the actual gene usage between different organisms may differ. However, given the teachings and guidance provided herein, those skilled in the art also will understand that the teachings and methods of the invention can be applied to all microbial organisms using the cognate metabolic alterations to those exemplified herein to construct a microbial organism in a species of interest that will synthesize alkene.

[0065] Methods for constructing and testing the expression levels of a non-naturally occurring alkene-producing host can be performed, for example, by recombinant and detection methods well known in the art. Such methods can be found described in, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Third Ed., Cold Spring Harbor Laboratory, New York (2001); and Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, MD (1999).

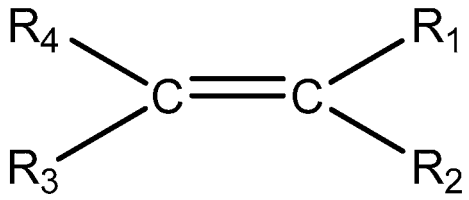
[0066] Exogenous nucleic acid sequences involved in a pathway for production of alkene can be introduced stably or transiently into a host cell using techniques well known in the art including, but not limited to, conjugation, electroporation, chemical transformation, transduction, transfection, and ultrasound transformation. For exogenous expression in *E. co*li or other prokaryotic cells, some nucleic acid sequences in the genes or cDNAs of eukaryotic nucleic acids can encode targeting signals such as an N-terminal mitochondrial or other targeting signal, which can be removed before transformation into prokaryotic host cells, if desired. For example, removal of a mitochondrial leader sequence led to increased expression in *E. coli* (Hoffmeister et al., *J. Biol. Chem.* 280:4329-4338 (2005)). For exogenous expression in yeast or other eukaryotic cells, genes can be expressed in the cytosol without the addition of leader sequence, or can be targeted to mitochondrion or other organelles, or targeted for secretion, by the addition of a suitable targeting sequence such as a mitochondrial targeting or secretion signal suitable for the host cells. Thus, it is understood that appropriate modifications to a nucleic acid sequence to remove or include a targeting sequence can be incorporated into an exogenous nucleic acid sequence to impart desirable properties. Furthermore, genes can be subjected

to codon optimization with techniques well known in the art to achieve optimized expression of the proteins.

[0067] An expression vector or vectors can be constructed to include one or more alkene biosynthetic pathway encoding nucleic acids as exemplified herein operably linked to expression control sequences functional in the host organism. Expression vectors applicable for use in the microbial host organisms of the invention include, for example, plasmids, phage vectors, viral vectors, episomes and artificial chromosomes, including vectors and selection sequences or markers operable for stable integration into a host chromosome. Additionally, the expression vectors can include one or more selectable marker genes and appropriate expression control sequences. Selectable marker genes also can be included that, for example, provide resistance to antibiotics or toxins, complement auxotrophic deficiencies, or supply critical nutrients not in the culture media. Expression control sequences can include constitutive and inducible promoters, transcription enhancers, transcription terminators, and the like which are well known in the art. When two or more exogenous encoding nucleic acids are to be co-expressed, both nucleic acids can be inserted, for example, into a single expression vector or in separate expression vectors. For single vector expression, the encoding nucleic acids can be operationally linked to one common expression control sequence or linked to different expression control sequences, such as one inducible promoter and one constitutive promoter. The transformation of exogenous nucleic acid sequences involved in a metabolic or synthetic pathway can be confirmed using methods well known in the art. Such methods include, for example, nucleic acid analysis such as Northern blots or polymerase chain reaction (PCR) amplification of mRNA, or immunoblotting for expression of gene products, or other suitable analytical methods to test the expression of an introduced nucleic acid sequence or its corresponding gene product. It is understood by those skilled in the art that the exogenous nucleic acid is expressed in a sufficient amount to produce the desired product, and it is further understood that expression levels can be optimized to obtain sufficient expression using methods well known in the art and as disclosed herein.

[0068] The invention additionally provides methods of using the microbial organisms disclosed herein to produce an alkene, by culturing a non-naturally occurring microbial organism containing an alkene pathway as described herein under conditions and for a sufficient period of time to produce an alkene. In some aspects of the method, the

microbial organism used in the method can produce an alkene, wherein the alkene is a compound of Formula (II)



[0069] wherein R^1, R^2, R^3, and R^4 are each independently (a) hydrogen, cyano, halo, or nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more substituents Q; or (c) $-C(O)R^{1a}$, $-C(O)OR^{1a}$, $-C(O)NR^{1b}R^{1c}$, $-C(NR^{1a})NR^{1b}R^{1c}$, $-OR^{1a}$, $-OC(O)R^{1a}$, $-OC(O)OR^{1a}$, $-OC(O)NR^{1b}R^{1c}$, $-OC(=NR^{1a})NR^{1b}R^{1c}$, $-OS(O)R^{1a}$, $-OS(O)_2R^{1a}$, $-OS(O)NR^{1b}R^{1c}$, $-OS(O)_2NR^{1b}R^{1c}$, $-NR^{1b}R^{1c}$, $-NR^{1a}C(O)R^{1d}$, $-NR^{1a}C(O)OR^{1d}$, $-NR^{1a}C(O)NR^{1b}R^{1c}$, $-NR^{1a}C(=NR^{1d})NR^{1b}R^{1c}$, $-NR^{1a}S(O)R^{1d}$, $-NR^{1a}S(O)_2R^{1d}$, $-NR^{1a}S(O)NR^{1b}R^{1c}$, $-NR^{1a}S(O)_2NR^{1b}R^{1c}$, $-SR^{1a}$, $-S(O)R^{1a}$, $-S(O)_2R^{1a}$, $-S(O)NR^{1b}R^{1c}$, or $-S(O)_2NR^{1b}R^{1c}$; wherein each R^{1a}, R^{1b}, R^{1c}, and R^{1d} is independently hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or R^{1a} and R^{1c} together with the C and N atoms to which they are attached form heterocyclyl; or R^{1b} and R^{1c} together with the N atom to which they are attached form heterocyclyl; wherein each Q is independently selected from (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl, each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (c) $-C(O)R^a$, $-C(O)OR^a$, $-C(O)NR^bR^c$, $-C(NR^a)NR^bR^c$, $-OR^a$, $-OC(O)R^a$, $-OC(O)OR^a$, $-OC(O)NR^bR^c$, $-OC(=NR^a)NR^bR^c$, $-OS(O)R^a$, $-OS(O)_2R^a$, $-OS(O)NR^bR^c$, $-OS(O)_2NR^bR^c$, $-NR^bR^c$, $-NR^aC(O)R^d$, $-NR^aC(O)OR^d$, $-NR^aC(O)NR^bR^c$, $-NR^aC(=NR^d)NR^bR^c$, $-NR^aS(O)R^d$, $-NR^aS(O)_2R^d$, $-NR^aS(O)NR^bR^c$, $-NR^aS(O)_2NR^bR^c$, $-SR^a$, $-S(O)R^a$, $-S(O)_2R^a$, $-S(O)NR^bR^c$, and $-S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heterocyclyl, optionally substituted with one or more, in one embodiment, one, two,

three, or four, substituents Q^a; wherein each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl; and (c) $-C(O)R^e$, $-C(O)OR^e$, $-C(O)NR^fR^g$, $-C(NR^e)NR^fR^g$, $-OR^e$, $-OC(O)R^e$, $-OC(O)OR^e$, $-OC(O)NR^fR^g$, $-OC(=NR^e)NR^fR^g$, $-OS(O)R^e$, $-OS(O)_2R^e$, $-OS(O)NR^fR^g$, $-OS(O)_2NR^fR^g$, $-NR^fR^g$, $-NR^eC(O)R^h$, $-NR^eC(O)OR^f$, $-NR^eC(O)NR^fR^g$, $-NR^eC(=NR^h)NR^fR^g$, $-NR^eS(O)R^h$, $-NR^eS(O)_2R^h$, $-NR^eS(O)NR^fR^g$, $-NR^eS(O)_2NR^fR^g$, $-SR^e$, $-S(O)R^e$, $-S(O)_2R^e$, $-S(O)NR^fR^g$, and $-S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heterocyclyl.

[0070] In some aspects, the microbial organism used in the method disclosed herein can produce an alkene, wherein the alkene is a compound selected from, but are not limited to, Ethylene, Propylene, Propylene, But-1-ene, Isobutylene, Isobutylene, But-1-ene, but-2-ene, Pent-1-ene, 3-methylbut-1-ene, Pent-2-ene, 2-methylbut-1-ene, 3-Methylbut-1-ene, 2-Methylbut-1-ene, 2-Methylbut-2-ene, Isoprene, 3-Methylbuta-1,2-diene, 1,3-Butadiene and Styrene.

[0071] In some embodiments the alkene product is gaseous and has limited solubility in the culture broth under the conditions of the process. This is advantageous, as removal of the gas from the reaction vessel can drive the alkene-forming pathway reactions in the forward direction.

[0072] Elevated temperature can further limit solubility of the alkene products. A desirable property of the microorganism containing the alkene-producing pathway is the ability to grow at elevated temperatures. Examplary thermophilic and heat-tolerant organisms include *Thermus aquaticus*, bacteria of the genus *Clostridium* and microorganisms of the genera *Thermotoga* and *Aquifex*. A desired property of the alkene-producing pathway enzymes is the ability to catalyze the desired reactions at elevated temperatures. Such enzymes can be isolated from thermophilic organisms or can be obtained by mutagenizing available enzymes and screening or selecting for increased activity under increased temperature conditions.

[0073] Suitable purification and/or assays to test for the production of an alkene can be performed using well known methods. Suitable replicates such as triplicate cultures can be grown for each engineered strain to be tested. For example, product and byproduct formation in the engineered production host can be monitored. The final product and intermediates, and other organic compounds, can be analyzed by methods such as HPLC (High Performance Liquid Chromatography), GC-MS (Gas Chromatography-Mass Spectroscopy) and LC-MS (Liquid Chromatography-Mass Spectroscopy) or other suitable analytical methods using routine procedures well known in the art. The release of product in the fermentation broth can also be tested with the culture supernatant. Byproducts and residual glucose can be quantified by HPLC using, for example, a refractive index detector for glucose and alcohols, and a UV detector for organic acids (Lin et al., *Biotechnol. Bioeng.* 90:775-779 (2005)), or other suitable assay and detection methods well known in the art. The individual enzyme or protein activities from the exogenous DNA sequences can also be assayed using methods well known in the art. Gaseous samples can be analyzed by gas chromatography (GC) coupled with a flame ionization detector, and further by GC-MS.

[0074] The alkene can be separated from other components in the culture using a variety of methods well known in the art. Such separation methods include, for example, extraction procedures as well as methods that include continuous liquid-liquid extraction, pervaporation, membrane filtration, membrane separation, reverse osmosis, electrodialysis, distillation, crystallization, centrifugation, extractive filtration, ion exchange chromatography, size exclusion chromatography, adsorption chromatography, and ultrafiltration. All of the above methods are well known in the art.

[0075] Any of the non-naturally occurring microbial organisms described herein can be cultured to produce and/or secrete the biosynthetic products of the invention. For example, the alkene producers can be cultured for the biosynthetic production of alkene.

[0076] For the production of alkene, the recombinant strains are cultured in a medium with carbon source and other essential nutrients. It is sometimes desirable and can be highly desirable to maintain anaerobic conditions in the fermenter to reduce the cost of the overall process. Such conditions can be obtained, for example, by first sparging the medium with nitrogen and then sealing the flasks with a septum and crimp-cap. For strains where growth is not observed anaerobically, microaerobic or substantially anaerobic conditions can be applied by perforating the septum with a small hole for

limited aeration. Exemplary anaerobic conditions have been described previously and are well-known in the art. Exemplary aerobic and anaerobic conditions are described, for example, in United State publication 2009/0047719, filed August 10, 2007. Fermentations can be performed in a batch, fed-batch or continuous manner, as disclosed herein.

[0077] If desired, the pH of the medium can be maintained at a desired pH, in particular neutral pH, such as a pH of around 7 by addition of a base, such as NaOH or other bases, or acid, as needed to maintain the culture medium at a desirable pH. The growth rate can be determined by measuring optical density using a spectrophotometer (600 nm), and the glucose uptake rate by monitoring carbon source depletion over time.

[0078] The growth medium can include, for example, any carbohydrate source which can supply a source of carbon to the non-naturally occurring microorganism. Such sources include, for example, sugars such as glucose, xylose, arabinose, galactose, mannose, fructose, sucrose and starch. Other sources of carbohydrate include, for example, renewable feedstocks and biomass. Exemplary types of biomasses that can be used as feedstocks in the methods of the invention include cellulosic biomass, hemicellulosic biomass and lignin feedstocks or portions of feedstocks. Such biomass feedstocks contain, for example, carbohydrate substrates useful as carbon sources such as glucose, xylose, arabinose, galactose, mannose, fructose and starch. Given the teachings and guidance provided herein, those skilled in the art will understand that renewable feedstocks and biomass other than those exemplified above also can be used for culturing the microbial organisms of the invention for the production of alkene.

[0079] In addition to renewable feedstocks such as those exemplified above, the alkene microbial organisms of the invention also can be modified for growth on syngas as its source of carbon. In this specific embodiment, one or more proteins or enzymes are expressed in the alkene producing organisms to provide a metabolic pathway for utilization of syngas or other gaseous carbon source.

[0080] Synthesis gas, also known as syngas or producer gas, is the major product of gasification of coal and of carbonaceous materials such as biomass materials, including agricultural crops and residues. Syngas is a mixture primarily of H_2 and CO and can be obtained from the gasification of any organic feedstock, including but not limited to coal, coal oil, natural gas, biomass, and waste organic matter. Gasification is generally carried out under a high fuel to oxygen ratio. Although largely H_2 and CO, syngas can also

include CO_2 and other gases in smaller quantities. Thus, synthesis gas provides a cost effective source of gaseous carbon such as CO and, additionally, CO_2.

[0081] The Wood-Ljungdahl pathway catalyzes the conversion of CO and H_2 to acetyl-CoA and other products such as acetate. Organisms capable of utilizing CO and syngas also generally have the capability of utilizing CO_2 and CO_2/H_2 mixtures through the same basic set of enzymes and transformations encompassed by the Wood-Ljungdahl pathway. H_2-dependent conversion of CO_2 to acetate by microorganisms was recognized long before it was revealed that CO also could be used by the same organisms and that the same pathways were involved. Many acetogens have been shown to grow in the presence of CO_2 and produce compounds such as acetate as long as hydrogen is present to supply the necessary reducing equivalents (see for example, Drake, *Acetogenesis*, pp. 3-60 Chapman and Hall, New York, (1994)). This can be summarized by the following equation:

$$2\ CO_2 + 4\ H_2 + n\ ADP + n\ Pi \rightarrow CH_3COOH + 2\ H_2O + n\ ATP$$

[0082] Hence, non-naturally occurring microorganisms possessing the Wood-Ljungdahl pathway can utilize CO_2 and H_2 mixtures as well for the production of acetyl-CoA and other desired products.

[0083] The Wood-Ljungdahl pathway is well known in the art and consists of 12 reactions which can be separated into two branches: (1) methyl branch and (2) carbonyl branch. The methyl branch converts syngas to methyl-tetrahydrofolate (methyl-THF) whereas the carbonyl branch converts methyl-THF to acetyl-CoA. The reactions in the methyl branch are catalyzed in order by the following enzymes or proteins: ferredoxin oxidoreductase, formate dehydrogenase, formyltetrahydrofolate synthetase, methenyltetrahydrofolate cyclodehydratase, methylenetetrahydrofolate dehydrogenase and methylenetetrahydrofolate reductase. The reactions in the carbonyl branch are catalyzed in order by the following enzymes or proteins: methyltetrahydrofolate:corrinoid protein methyltransferase (for example, AcsE), corrinoid iron-sulfur protein, nickel-protein assembly protein (for example, AcsF), ferredoxin, acetyl-CoA synthase, carbon monoxide dehydrogenase and nickel-protein assembly protein (for example, CooC). Following the teachings and guidance provided herein for introducing a sufficient number of encoding nucleic acids to generate an alkene pathway, those skilled in the art will understand that the same engineering design also can be performed with respect to introducing at least the

nucleic acids encoding the Wood-Ljungdahl enzymes or proteins absent in the host organism. Therefore, introduction of one or more encoding nucleic acids into the microbial organisms of the invention such that the modified organism contains the complete Wood-Ljungdahl pathway will confer syngas utilization ability.

[0084] Additionally, the reductive (reverse) tricarboxylic acid cycle coupled with carbon monoxide dehydrogenase and/or hydrogenase activities can also be used for the conversion of CO, CO_2 and/or H_2 to acetyl-CoA and other products such as acetate. Organisms capable of fixing carbon via the reductive TCA pathway can utilize one or more of the following enzymes: ATP citrate-lyase, citrate lyase, aconitase, isocitrate dehydrogenase, alpha-ketoglutarate:ferredoxin oxidoreductase, succinyl-CoA synthetase, succinyl-CoA transferase, fumarate reductase, fumarase, malate dehydrogenase, NAD(P)H:ferredoxin oxidoreductase, carbon monoxide dehydrogenase, and hydrogenase. Specifically, the reducing equivalents extracted from CO and/or H_2 by carbon monoxide dehydrogenase and hydrogenase are utilized to fix CO_2 via the reductive TCA cycle into acetyl-CoA or acetate. Acetate can be converted to acetyl-CoA by enzymes such as acetyl-CoA transferase, acetate kinase/phosphotransacetylase, and acetyl-CoA synthetase. Acetyl-CoA can be converted to the alkene precursors, glyceraldehyde-3-phosphate, phosphoenolpyruvate, and pyruvate, by pyruvate:ferredoxin oxidoreductase and the enzymes of gluconeogenesis. Following the teachings and guidance provided herein for introducing a sufficient number of encoding nucleic acids to generate an alkene pathway, those skilled in the art will understand that the same engineering design also can be performed with respect to introducing at least the nucleic acids encoding the reductive TCA pathway enzymes or proteins absent in the host organism. Therefore, introduction of one or more encoding nucleic acids into the microbial organisms of the invention such that the modified organism contains a reductive TCA pathway can confer syngas utilization ability.

[0085] Accordingly, given the teachings and guidance provided herein, those skilled in the art will understand that a non-naturally occurring microbial organism can be produced that secretes the biosynthesized compounds of the invention when grown on a carbon source such as a carbohydrate. Such compounds include, for example, alkene and any of the intermediate metabolites in the alkene pathway. All that is required is to engineer in one or more of the required enzyme or protein activities to achieve biosynthesis of the desired compound or intermediate including, for example, inclusion of some or all of the

alkene biosynthetic pathways. Accordingly, the invention provides a non-naturally occurring microbial organism that produces and/or secretes alkene when grown on a carbohydrate or other carbon source and produces and/or secretes any of the intermediate metabolites shown in the alkene pathway when grown on a carbohydrate or other carbon source. The alkene producing microbial organisms of the invention can initiate synthesis from an intermediate, for example, an alkyl phosphate or an alkyl diphosphate.

[0086] The non-naturally occurring microbial organisms of the invention are constructed using methods well known in the art as exemplified herein to exogenously express at least one nucleic acid encoding an alkene pathway enzyme or protein in sufficient amounts to produce alkene. It is understood that the microbial organisms of the invention are cultured under conditions sufficient to produce alkene. Following the teachings and guidance provided herein, the non-naturally occurring microbial organisms of the invention can achieve biosynthesis of alkene resulting in intracellular concentrations between about 0.1-200 mM or more. Generally, the intracellular concentration of alkene is between about 3-150 mM, particularly between about 5-125 mM and more particularly between about 8-100 mM, including about 10 mM, 20 mM, 50 mM, 80 mM, or more. Intracellular concentrations between and above each of these exemplary ranges also can be achieved from the non-naturally occurring microbial organisms of the invention.

[0087] In some embodiments, culture conditions include anaerobic or substantially anaerobic growth or maintenance conditions. Exemplary anaerobic conditions have been described previously and are well known in the art. Exemplary anaerobic conditions for fermentation processes are described herein and are described, for example, in U.S. publication 2009/0047719, filed August 10, 2007. Any of these conditions can be employed with the non-naturally occurring microbial organisms as well as other anaerobic conditions well known in the art. Under such anaerobic or substantially anaerobic conditions, the alkene producers can synthesize alkene at intracellular concentrations of 5-10 mM or more as well as all other concentrations exemplified herein. It is understood that, even though the above description refers to intracellular concentrations, alkene producing microbial organisms can produce alkene intracellularly and/or secrete the product into the culture medium.

[0088] In addition to the culturing and fermentation conditions disclosed herein, growth condition for achieving biosynthesis of alkene can include the addition of an osmoprotectant to the culturing conditions. In certain embodiments, the non-naturally

occurring microbial organisms of the invention can be sustained, cultured or fermented as described herein in the presence of an osmoprotectant. Briefly, an osmoprotectant refers to a compound that acts as an osmolyte and helps a microbial organism as described herein survive osmotic stress. Osmoprotectants include, but are not limited to, betaines, amino acids, and the sugar trehalose. Non-limiting examples of such are glycine betaine, praline betaine, dimethylthetin, dimethylslfonioproprionate, 3-dimethylsulfonio-2-methylproprionate, pipecolic acid, dimethylsulfonioacetate, choline, L-carnitine and ectoine. In one aspect, the osmoprotectant is glycine betaine. It is understood to one of ordinary skill in the art that the amount and type of osmoprotectant suitable for protecting a microbial organism described herein from osmotic stress will depend on the microbial organism used. The amount of osmoprotectant in the culturing conditions can be, for example, no more than about 0.1 mM, no more than about 0.5 mM, no more than about 1.0 mM, no more than about 1.5 mM, no more than about 2.0 mM, no more than about 2.5 mM, no more than about 3.0 mM, no more than about 5.0 mM, no more than about 7.0 mM, no more than about 10mM, no more than about 50mM, no more than about 100mM or no more than about 500mM.

[0089] In some embodiments, the carbon feedstock and other cellular uptake sources such as phosphate, ammonia, sulfate, chloride and other halogens can be chosen to alter the isotopic distribution of the atoms present in an alkene or any alkene pathway intermediate. The various carbon feedstock and other uptake sources enumerated above will be referred to herein, collectively, as "uptake sources." Uptake sources can provide isotopic enrichment for any atom present in the product alkene or alkene pathway intermediate, or for side products generated in reactions diverging away from an alkene pathway. Isotopic enrichment can be achieved for any target atom including, for example, carbon, hydrogen, oxygen, nitrogen, sulfur, phosphorus, chloride or other halogens.

[0090] In some embodiments, the uptake sources can be selected to alter the carbon-12, carbon-13, and carbon-14 ratios. In some embodiments, the uptake sources can be selected to alter the oxygen-16, oxygen-17, and oxygen-18 ratios. In some embodiments, the uptake sources can be selected to alter the hydrogen, deuterium, and tritium ratios. In some embodiments, the uptake sources can be selected to alter the nitrogen-14 and nitrogen-15 ratios. In some embodiments, the uptake sources can be selected to alter the sulfur-32, sulfur-33, sulfur-34, and sulfur-35 ratios. In some embodiments, the uptake sources can be selected to alter the phosphorus-31, phosphorus-32, and phosphorus-33

ratios. In some embodiments, the uptake sources can be selected to alter the chlorine-35, chlorine-36, and chlorine-37 ratios.

[0091] In some embodiments, the isotopic ratio of a target atom can be varied to a desired ratio by selecting one or more uptake sources. An uptake source can be derived from a natural source, as found in nature, or from a man-made source, and one skilled in the art can select a natural source, a man-made source, or a combination thereof, to achieve a desired isotopic ratio of a target atom. An example of a man-made uptake source includes, for example, an uptake source that is at least partially derived from a chemical synthetic reaction. Such isotopically enriched uptake sources can be purchased commercially or prepared in the laboratory and/or optionally mixed with a natural source of the uptake source to achieve a desired isotopic ratio. In some embodiments, a target atom isotopic ratio of an uptake source can be achieved by selecting a desired origin of the uptake source as found in nature. For example, as discussed herein, a natural source can be a biobased derived from or synthesized by a biological organism or a source such as petroleum-based products or the atmosphere. In some such embodiments, a source of carbon, for example, can be selected from a fossil fuel-derived carbon source, which can be relatively depleted of carbon-14, or an environmental or atmospheric carbon source, such as CO_2, which can possess a larger amount of carbon-14 than its petroleum-derived counterpart.

[0092] The unstable carbon isotope carbon-14 or radiocarbon makes up for roughly 1 in 10^{12} carbon atoms in the earth's atmosphere and has a half-life of about 5700 years. The stock of carbon is replenished in the upper atmosphere by a nuclear reaction involving cosmic rays and ordinary nitrogen (^{14}N). Fossil fuels contain no carbon-14, as it decayed long ago. Burning of fossil fuels lowers the atmospheric carbon-14 fraction, the so-called "Suess effect".

[0093] Methods of determining the isotopic ratios of atoms in a compound are well known to those skilled in the art. Isotopic enrichment is readily assessed by mass spectrometry using techniques known in the art such as accelerated mass spectrometry (AMS), Stable Isotope Ratio Mass Spectrometry (SIRMS) and Site-Specific Natural Isotopic Fractionation by Nuclear Magnetic Resonance (SNIF-NMR). Such mass spectral techniques can be integrated with separation techniques such as liquid chromatography (LC), high performance liquid chromatography (HPLC) and/or gas chromatography, and the like.

[0094] In the case of carbon, ASTM D6866 was developed in the United States as a standardized analytical method for determining the biobased content of solid, liquid, and gaseous samples using radiocarbon dating by the American Society for Testing and Materials (ASTM) International. The standard is based on the use of radiocarbon dating for the determination of a product's biobased content. ASTM D6866 was first published in 2004, and the current active version of the standard is ASTM D6866-11 (effective April 1, 2011). Radiocarbon dating techniques are well known to those skilled in the art, including those described herein.

[0095] The biobased content of a compound is estimated by the ratio of carbon-14 (^{14}C) to carbon-12 (^{12}C). Specifically, the Fraction Modern (Fm) is computed from the expression: Fm = (S-B)/(M-B), where B, S and M represent the $^{14}C/^{12}C$ ratios of the blank, the sample and the modern reference, respectively. Fraction Modern is a measurement of the deviation of the $^{14}C/^{12}C$ ratio of a sample from "Modern." Modern is defined as 95% of the radiocarbon concentration (in AD 1950) of National Bureau of Standards (NBS) Oxalic Acid I (i.e., standard reference materials (SRM) 4990b) normalized to $\delta^{13}C_{VPDB}$=-19 per mil (Olsson, *The use of Oxalic acid as a Standard.* in, Radiocarbon Variations and Absolute Chronology, Nobel Symposium, 12th Proc., John Wiley & Sons, New York (1970)). Mass spectrometry results, for example, measured by ASM, are calculated using the internationally agreed upon definition of 0.95 times the specific activity of NBS Oxalic Acid I (SRM 4990b) normalized to $\delta^{13}C_{VPDB}$=-19 per mil. This is equivalent to an absolute (AD 1950) $^{14}C/^{12}C$ ratio of $1.176 \pm 0.010 \times 10^{-12}$ (Karlen et al., *Arkiv Geofysik*, 4:465-471 (1968)). The standard calculations take into account the differential uptake of one istope with respect to another, for example, the preferential uptake in biological systems of C^{12} over C^{13} over C^{14}, and these corrections are reflected as a Fm corrected for δ^{13}.

[0096] An oxalic acid standard (SRM 4990b or HOx 1) was made from a crop of 1955 sugar beet. Although there were 1000 lbs made, this oxalic acid standard is no longer commercially available. The Oxalic Acid II standard (HOx 2; N.I.S.T designation SRM 4990 C) was made from a crop of 1977 French beet molasses. In the early 1980's, a group of 12 laboratories measured the ratios of the two standards. The ratio of the activity of Oxalic acid II to 1 is 1.2933±0.001 (the weighted mean). The isotopic ratio of HOx II is -17.8 per mille. ASTM D6866-11 suggests use of the available Oxalic Acid II standard SRM 4990 C (Hox2) for the modern standard (see discussion of original vs. currently

available oxalic acid standards in Mann, *Radiocarbon*, 25(2):519-527 (1983)). A Fm = 0% represents the entire lack of carbon-14 atoms in a material, thus indicating a fossil (for example, petroleum based) carbon source. A Fm = 100%, after correction for the post-1950 injection of carbon-14 into the atmosphere from nuclear bomb testing, indicates an entirely modern carbon source. As described herein, such a "modern" source includes biobased sources.

[0097] As described in ASTM D6866, the percent modern carbon (pMC) can be greater than 100% because of the continuing but diminishing effects of the 1950s nuclear testing programs, which resulted in a considerable enrichment of carbon-14 in the atmosphere as described in ASTM D6866-11. Because all sample carbon-14 activities are referenced to a "pre-bomb" standard, and because nearly all new biobased products are produced in a post-bomb environment, all pMC values (after correction for isotopic fraction) must be multiplied by 0.95 (as of 2010) to better reflect the true biobased content of the sample. A biobased content that is greater than 103% suggests that either an analytical error has occurred, or that the source of biobased carbon is more than several years old.

[0098] ASTM D6866 quantifies the biobased content relative to the material's total organic content and does not consider the inorganic carbon and other non-carbon containing substances present. For example, a product that is 50% starch-based material and 50% water would be considered to have a Biobased Content = 100% (50% organic content that is 100% biobased) based on ASTM D6866. In another example, a product that is 50% starch-based material, 25% petroleum-based, and 25% water would have a Biobased Content = 66.7% (75% organic content but only 50% of the product is biobased). In another example, a product that is 50% organic carbon and is a petroleum-based product would be considered to have a Biobased Content = 0% (50% organic carbon but from fossil sources). Thus, based on the well known methods and known standards for determining the biobased content of a compound or material, one skilled in the art can readily determine the biobased content and/or prepared downstream products that utilize of the invention having a desired biobased content.

[0099] Applications of carbon-14 dating techniques to quantify bio-based content of materials are known in the art (Currie et al., *Nuclear Instruments and Methods in Physics Research B,* 172:281-287 (2000)). For example, carbon-14 dating has been used to quantify bio-based content in terephthalate-containing materials (Colonna et al., *Green*

Chemistry, 13:2543-2548 (2011)). Notably, polypropylene terephthalate (PPT) polymers derived from renewable 1,3-propanediol and petroleum-derived terephthalic acid resulted in Fm values near 30% (i.e., since 3/11 of the polymeric carbon derives from renewable 1,3-propanediol and 8/11 from the fossil end member terephthalic acid) (Currie et al., *supra*, 2000). In contrast, polybutylene terephthalate polymer derived from both renewable 1,4-butanediol and renewable terephthalic acid resulted in bio-based content exceeding 90% (Colonna et al., *supra*, 2011).

[00100] Accordingly, in some embodiments, the present invention provides an alkene or an alkene pathway intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that reflects an atmospheric carbon, also referred to as environmental carbon, uptake source. For example, in some aspects the alkene or alkene pathway intermediate can have an Fm value of at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or as much as 100%. In some such embodiments, the uptake source is CO_2. In some embodiments, the present invention provides an alkene or an alkene pathway intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that reflects petroleum-based carbon uptake source. In this aspect, the alkene or alkene pathway intermediate can have an Fm value of less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, less than 50%, less than 45%, less than 40%, less than 35%, less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2% or less than 1%. In some embodiments, the present invention provides an alkene or an alkene pathway intermediate that has a carbon-12, carbon-13, and carbon-14 ratio that is obtained by a combination of an atmospheric carbon uptake source with a petroleum-based uptake source. Using such a combination of uptake sources is one way by which the carbon-12, carbon-13, and carbon-14 ratio can be varied, and the respective ratios would reflect the proportions of the uptake sources.

[00101] Further, the present invention relates to the biologically produced alkene or alkene pathway intermediate as disclosed herein, and to the products derived therefrom, wherein the alkene or alkene pathway intermediate has a carbon-12, carbon-13, and carbon-14 isotope ratio of about the same value as the CO_2 that occurs in the environment. For example, in some aspects the invention provides bioderived alkene or a bioderived alkene intermediate having a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of

about the same value as the CO_2 that occurs in the environment, or any of the other ratios disclosed herein. It is understood, as disclosed herein, that a product can have a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of about the same value as the CO_2 that occurs in the environment, or any of the ratios disclosed herein, wherein the product is generated from a bioderived alkene or a bioderived alkene pathway intermediate as disclosed herein, wherein the bioderived product is chemically modified to generate a final product. Methods of chemically modifying a bioderived product of alkene, or an intermediate thereof, to generate a desired product are well known to those skilled in the art, as described herein. The invention further provides a plastic, a polymer, a co-polymer, a polymer intermediate, a resin, a rubber, or a fiber having a carbon-12 versus carbon-13 versus carbon-14 isotope ratio of about the same value as the CO_2 that occurs in the environment, wherein the a plastic, a polymer, a co-polymer, a polymer intermediate, a resin, a rubber, or a fiber are generated directly from or in combination with bioderived alkene or a bioderived alkene pathway intermediate as disclosed herein.

[00102] Alkenes include a variety chemicals as described herein, which can be used in commercial and industrial applications. For example, the alkenes disclosed herein can be used as a raw material in the production of a wide range of products including plastics, polymers, co-polymers, polymer intermediates, resins, rubbers, or fibers. Accordingly, in some embodiments, the invention provides biobased plastics, polymers, co-polymers, polymer intermediates, resins, rubbers, or fibers comprising one or more bioderived alkene or bioderived alkene pathway intermediate produced by a non-naturally occurring microorganism of the invention or produced using a method disclosed herein.

[00103] As used herein, the term "bioderived" means derived from or synthesized by a biological organism and can be considered a renewable resource since it can be generated by a biological organism. Such a biological organism, in particular the microbial organisms of the invention disclosed herein, can utilize feedstock or biomass, such as, sugars or carbohydrates obtained from an agricultural, plant, bacterial, or animal source. Alternatively, the biological organism can utilize atmospheric carbon. As used herein, the term "biobased" means a product as described above that is composed, in whole or in part, of a bioderived compound of the invention. A biobased or bioderived product is in contrast to a petroleum derived product, wherein such a product is derived from or synthesized from petroleum or a petrochemical feedstock.

[00104] In some embodiments, the invention provides a plastic, a polymer, a co-polymer, a polymer intermediate, a resin, a rubber, or a fiber comprising a bioderived alkene or a bioderived alkene pathway intermediate, wherein the bioderived alkene or bioderived alkene pathway intermediate includes all or part of the alkene or alkene pathway intermediate used in the production of the plastic, polymer, co-polymer, polymer intermediate, resin, rubber, or fiber. Thus, in some aspects, the invention provides a biobased plastic, polymer, co-polymer, polymer intermediate, resin, rubber, or fiber comprising at least 2%, at least 3%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98% or 100% bioderived alkene or bioderived alkene pathway intermediate as disclosed herein. Additionally, in some aspects, the invention provides a biobased plastic, polymer, co-polymer, polymer intermediate, resin, rubber, or fiber wherein the alkene or alkene pathway intermediate used in its production is a combination of bioderived and petroleum derived alkene or alkene pathway intermediate. For example, a biobased plastic, polymer, co-polymer, polymer intermediate, resin, rubber, or fiber can be produced using 50% bioderived alkene and 50% petroleum derived alkene or other desired ratios such as 60%/40%, 70%/30%, 80%/20%, 90%/10%, 95%/5%, 100%/0%, 40%/60%, 30%/70%, 20%/80%, 10%/90% of bioderived/petroleum derived precursors, so long as at least a portion of the product comprises a bioderived product produced by the microbial organisms disclosed herein. It is understood that methods for producing plastic, polymer, co-polymer, polymer intermediate, resin, rubber, or fiber using the bioderived alkene or bioderived alkene pathway intermediate of the invention are well known in the art.

[00105] The culture conditions can include, for example, liquid culture procedures as well as fermentation and other large scale culture procedures. As described herein, particularly useful yields of the biosynthetic products of the invention can be obtained under anaerobic or substantially anaerobic culture conditions.

[00106] As described herein, one exemplary growth condition for achieving biosynthesis of alkene includes anaerobic culture or fermentation conditions. In certain embodiments, the non-naturally occurring microbial organisms of the invention can be sustained, cultured or fermented under anaerobic or substantially anaerobic conditions. Briefly, anaerobic conditions refers to an environment devoid of oxygen. Substantially anaerobic conditions include, for example, a culture, batch fermentation or continuous

fermentation such that the dissolved oxygen concentration in the medium remains between 0 and 10% of saturation. Substantially anaerobic conditions also includes growing or resting cells in liquid medium or on solid agar inside a sealed chamber maintained with an atmosphere of less than 1% oxygen. The percent of oxygen can be maintained by, for example, sparging the culture with an N_2/CO_2 mixture or other suitable non-oxygen gas or gases.

[00107] The culture conditions described herein can be scaled up and grown continuously for manufacturing of alkene. Exemplary growth procedures include, for example, fed-batch fermentation and batch separation; fed-batch fermentation and continuous separation, or continuous fermentation and continuous separation. All of these processes are well known in the art. Fermentation procedures are particularly useful for the biosynthetic production of commercial quantities of alkene. Generally, and as with non-continuous culture procedures, the continuous and/or near-continuous production of alkene will include culturing a non-naturally occurring alkene producing organism of the invention in sufficient nutrients and medium to sustain and/or nearly sustain growth in an exponential phase. Continuous culture under such conditions can include, for example, growth for 1 day, 2, 3, 4, 5, 6 or 7 days or more. Additionally, continuous culture can include longer time periods of 1 week, 2, 3, 4 or 5 or more weeks and up to several months. Alternatively, organisms of the invention can be cultured for hours, if suitable for a particular application. It is to be understood that the continuous and/or near-continuous culture conditions also can include all time intervals in between these exemplary periods. It is further understood that the time of culturing the microbial organism of the invention is for a sufficient period of time to produce a sufficient amount of product for a desired purpose.

[00108] Fermentation procedures are well known in the art. Briefly, fermentation for the biosynthetic production of alkene can be utilized in, for example, fed-batch fermentation and batch separation; fed-batch fermentation and continuous separation, or continuous fermentation and continuous separation. Examples of batch and continuous fermentation procedures are well known in the art.

[00109] In addition to the above fermentation procedures using the alkene producers of the invention for continuous production of substantial quantities of alkene, the alkene producers also can be, for example, simultaneously subjected to chemical synthesis procedures to convert the product to other compounds or the product can be separated

from the fermentation culture and sequentially subjected to chemical or enzymatic conversion to convert the product to other compounds, if desired.

[00110] To generate better producers, metabolic modeling can be utilized to optimize growth conditions. Modeling can also be used to design gene knockouts that additionally optimize utilization of the pathway (see, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and U.S. Patent No. 7,127,379). Modeling analysis allows reliable predictions of the effects on cell growth of shifting the metabolism towards more efficient production of alkene.

[00111] One computational method for identifying and designing metabolic alterations favoring biosynthesis of a desired product is the OptKnock computational framework (Burgard et al., *Biotechnol. Bioeng*. 84:647-657 (2003)). OptKnock is a metabolic modeling and simulation program that suggests gene deletion or disruption strategies that result in genetically stable microorganisms which overproduce the target product. Specifically, the framework examines the complete metabolic and/or biochemical network of a microorganism in order to suggest genetic manipulations that force the desired biochemical to become an obligatory byproduct of cell growth. By coupling biochemical production with cell growth through strategically placed gene deletions or other functional gene disruption, the growth selection pressures imposed on the engineered strains after long periods of time in a bioreactor lead to improvements in performance as a result of the compulsory growth-coupled biochemical production. Lastly, when gene deletions are constructed there is a negligible possibility of the designed strains reverting to their wild-type states because the genes selected by OptKnock are to be completely removed from the genome. Therefore, this computational methodology can be used to either identify alternative pathways that lead to biosynthesis of a desired product or used in connection with the non-naturally occurring microbial organisms for further optimization of biosynthesis of a desired product.

[00112] Briefly, OptKnock is a term used herein to refer to a computational method and system for modeling cellular metabolism. The OptKnock program relates to a framework of models and methods that incorporate particular constraints into flux balance analysis (FBA) models. These constraints include, for example, qualitative kinetic information, qualitative regulatory information, and/or DNA microarray experimental data. OptKnock also computes solutions to various metabolic problems by, for example, tightening the flux

boundaries derived through flux balance models and subsequently probing the performance limits of metabolic networks in the presence of gene additions or deletions. OptKnock computational framework allows the construction of model formulations that allow an effective query of the performance limits of metabolic networks and provides methods for solving the resulting mixed-integer linear programming problems. The metabolic modeling and simulation methods referred to herein as OptKnock are described in, for example, U.S. publication 2002/0168654, filed January 10, 2002, in International Patent No. PCT/US02/00660, filed January 10, 2002, and U.S. publication 2009/0047719, filed August 10, 2007.

[00113] Another computational method for identifying and designing metabolic alterations favoring biosynthetic production of a product is a metabolic modeling and simulation system termed SimPheny®. This computational method and system is described in, for example, U.S. publication 2003/0233218, filed June 14, 2002, and in International Patent Application No. PCT/US03/18838, filed June 13, 2003. SimPheny® is a computational system that can be used to produce a network model *in silico* and to simulate the flux of mass, energy or charge through the chemical reactions of a biological system to define a solution space that contains any and all possible functionalities of the chemical reactions in the system, thereby determining a range of allowed activities for the biological system. This approach is referred to as constraints-based modeling because the solution space is defined by constraints such as the known stoichiometry of the included reactions as well as reaction thermodynamic and capacity constraints associated with maximum fluxes through reactions. The space defined by these constraints can be interrogated to determine the phenotypic capabilities and behavior of the biological system or of its biochemical components.

[00114] These computational approaches are consistent with biological realities because biological systems are flexible and can reach the same result in many different ways. Biological systems are designed through evolutionary mechanisms that have been restricted by fundamental constraints that all living systems must face. Therefore, constraints-based modeling strategy embraces these general realities. Further, the ability to continuously impose further restrictions on a network model via the tightening of constraints results in a reduction in the size of the solution space, thereby enhancing the precision with which physiological performance or phenotype can be predicted.

[00115] Given the teachings and guidance provided herein, those skilled in the art will be able to apply various computational frameworks for metabolic modeling and simulation to design and implement biosynthesis of a desired compound in host microbial organisms. Such metabolic modeling and simulation methods include, for example, the computational systems exemplified above as SimPheny® and OptKnock. For illustration of the invention, some methods are described herein with reference to the OptKnock computation framework for modeling and simulation. Those skilled in the art will know how to apply the identification, design and implementation of the metabolic alterations using OptKnock to any of such other metabolic modeling and simulation computational frameworks and methods well known in the art.

[00116] The methods described above will provide one set of metabolic reactions to disrupt. Elimination of each reaction within the set or metabolic modification can result in a desired product as an obligatory product during the growth phase of the organism. Because the reactions are known, a solution to the bilevel OptKnock problem also will provide the associated gene or genes encoding one or more enzymes that catalyze each reaction within the set of reactions. Identification of a set of reactions and their corresponding genes encoding the enzymes participating in each reaction is generally an automated process, accomplished through correlation of the reactions with a reaction database having a relationship between enzymes and encoding genes.

[00117] Once identified, the set of reactions that are to be disrupted in order to achieve production of a desired product are implemented in the target cell or organism by functional disruption of at least one gene encoding each metabolic reaction within the set. One particularly useful means to achieve functional disruption of the reaction set is by deletion of each encoding gene. However, in some instances, it can be beneficial to disrupt the reaction by other genetic aberrations including, for example, mutation, deletion of regulatory regions such as promoters or cis binding sites for regulatory factors, or by truncation of the coding sequence at any of a number of locations. These latter aberrations, resulting in less than total deletion of the gene set can be useful, for example, when rapid assessments of the coupling of a product are desired or when genetic reversion is less likely to occur.

[00118] To identify additional productive solutions to the above described bilevel OptKnock problem which lead to further sets of reactions to disrupt or metabolic modifications that can result in the biosynthesis, including growth-coupled biosynthesis of

a desired product, an optimization method, termed integer cuts, can be implemented. This method proceeds by iteratively solving the OptKnock problem exemplified above with the incorporation of an additional constraint referred to as an integer cut at each iteration. Integer cut constraints effectively prevent the solution procedure from choosing the exact same set of reactions identified in any previous iteration that obligatorily couples product biosynthesis to growth. For example, if a previously identified growth-coupled metabolic modification specifies reactions 1, 2, and 3 for disruption, then the following constraint prevents the same reactions from being simultaneously considered in subsequent solutions. The integer cut method is well known in the art and can be found described in, for example, Burgard et al., *Biotechnol. Prog.* 17:791-797 (2001). As with all methods described herein with reference to their use in combination with the OptKnock computational framework for metabolic modeling and simulation, the integer cut method of reducing redundancy in iterative computational analysis also can be applied with other computational frameworks well known in the art including, for example, SimPheny®.

[00119] The methods exemplified herein allow the construction of cells and organisms that biosynthetically produce a desired product, including the obligatory coupling of production of a target biochemical product to growth of the cell or organism engineered to harbor the identified genetic alterations. Therefore, the computational methods described herein allow the identification and implementation of metabolic modifications that are identified by an *in silico* method selected from OptKnock or SimPheny®. The set of metabolic modifications can include, for example, addition of one or more biosynthetic pathway enzymes and/or functional disruption of one or more metabolic reactions including, for example, disruption by gene deletion.

[00120] As discussed above, the OptKnock methodology was developed on the premise that mutant microbial networks can be evolved towards their computationally predicted maximum-growth phenotypes when subjected to long periods of growth selection. In other words, the approach leverages an organism's ability to self-optimize under selective pressures. The OptKnock framework allows for the exhaustive enumeration of gene deletion combinations that force a coupling between biochemical production and cell growth based on network stoichiometry. The identification of optimal gene/reaction knockouts requires the solution of a bilevel optimization problem that chooses the set of active reactions such that an optimal growth solution for the resulting network

overproduces the biochemical of interest (Burgard et al., *Biotechnol. Bioeng.* 84:647-657 (2003)).

[00121] An *in silico* stoichiometric model of *E. coli* metabolism can be employed to identify essential genes for metabolic pathways as exemplified previously and described in, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and in U.S. Patent No. 7,127,379. As disclosed herein, the OptKnock mathematical framework can be applied to pinpoint gene deletions leading to the growth-coupled production of a desired product. Further, the solution of the bilevel OptKnock problem provides only one set of deletions. To enumerate all meaningful solutions, that is, all sets of knockouts leading to growth-coupled production formation, an optimization technique, termed integer cuts, can be implemented. This entails iteratively solving the OptKnock problem with the incorporation of an additional constraint referred to as an integer cut at each iteration, as discussed above.

[00122] As disclosed herein, a nucleic acid encoding a desired activity of an alkene pathway can be introduced into a host organism. In some cases, it can be desirable to modify an activity of an alkene pathway enzyme or protein to increase production of alkene. For example, known mutations that increase the activity of a protein or enzyme can be introduced into an encoding nucleic acid molecule. Additionally, optimization methods can be applied to increase the activity of an enzyme or protein and/or decrease an inhibitory activity, for example, decrease the activity of a negative regulator.

[00123] One such optimization method is directed evolution. Directed evolution is a powerful approach that involves the introduction of mutations targeted to a specific gene in order to improve and/or alter the properties of an enzyme. Improved and/or altered enzymes can be identified through the development and implementation of sensitive high-throughput screening assays that allow the automated screening of many enzyme variants (for example, $>10^4$). Iterative rounds of mutagenesis and screening typically are performed to afford an enzyme with optimized properties. Computational algorithms that can help to identify areas of the gene for mutagenesis also have been developed and can significantly reduce the number of enzyme variants that need to be generated and screened. Numerous directed evolution technologies have been developed (for reviews, see Hibbert et al., *Biomol.Eng* 22:11-19 (2005); Huisman and Lalonde, *In* Biocatalysis in the pharmaceutical and biotechnology industries pgs. 717-742 (2007), Patel (ed.), CRC

Press; Otten and Quax. *Biomol.Eng* 22:1-9 (2005).; and Sen et al., *Appl Biochem.Biotechnol* 143:212-223 (2007)) to be effective at creating diverse variant libraries, and these methods have been successfully applied to the improvement of a wide range of properties across many enzyme classes. Enzyme characteristics that have been improved and/or altered by directed evolution technologies include, for example: selectivity/specificity, for conversion of non-natural substrates; temperature stability, for robust high temperature processing; pH stability, for bioprocessing under lower or higher pH conditions; substrate or product tolerance, so that high product titers can be achieved; binding (K_m), including broadening substrate binding to include non-natural substrates; inhibition (K_i), to remove inhibition by products, substrates, or key intermediates; activity (kcat), to increases enzymatic reaction rates to achieve desired flux; expression levels, to increase protein yields and overall pathway flux; oxygen stability, for operation of air sensitive enzymes under aerobic conditions; and anaerobic activity, for operation of an aerobic enzyme in the absence of oxygen.

[00124] A number of exemplary methods have been developed for the mutagenesis and diversification of genes to target desired properties of specific enzymes. Such methods are well known to those skilled in the art. Any of these can be used to alter and/or optimize the activity of an alkene pathway enzyme or protein. Such methods include, but are not limited to EpPCR, which introduces random point mutations by reducing the fidelity of DNA polymerase in PCR reactions (Pritchard et al., *J Theor.Biol.* 234:497-509 (2005)); Error-prone Rolling Circle Amplification (epRCA), which is similar to epPCR except a whole circular plasmid is used as the template and random 6-mers with exonuclease resistant thiophosphate linkages on the last 2 nucleotides are used to amplify the plasmid followed by transformation into cells in which the plasmid is re-circularized at tandem repeats (Fujii et al., *Nucleic Acids Res.* 32:e145 (2004); and Fujii et al., *Nat. Protoc.* 1:2493-2497 (2006)); DNA or Family Shuffling, which typically involves digestion of two or more variant genes with nucleases such as Dnase I or EndoV to generate a pool of random fragments that are reassembled by cycles of annealing and extension in the presence of DNA polymerase to create a library of chimeric genes (Stemmer, *Proc Natl Acad Sci USA* 91:10747-10751 (1994); and Stemmer, *Nature* 370:389-391 (1994)); Staggered Extension (StEP), which entails template priming followed by repeated cycles of 2 step PCR with denaturation and very short duration of annealing/extension (as short as 5 sec) (Zhao et al., *Nat. Biotechnol.* 16:258-261 (1998)); Random Priming Recombination (RPR), in which random sequence primers are used to generate many short

DNA fragments complementary to different segments of the template (Shao et al., *Nucleic Acids Res* 26:681-683 (1998)).

[00125] Additional methods include Heteroduplex Recombination, in which linearized plasmid DNA is used to form heteroduplexes that are repaired by mismatch repair (Volkov et al, *Nucleic Acids Res.* 27:e18 (1999); and Volkov et al., *Methods Enzymol.* 328:456-463 (2000)); Random Chimeragenesis on Transient Templates (RACHITT), which employs Dnase I fragmentation and size fractionation of single stranded DNA (ssDNA) (Coco et al., *Nat. Biotechnol.* 19:354-359 (2001)); Recombined Extension on Truncated templates (RETT), which entails template switching of unidirectionally growing strands from primers in the presence of unidirectional ssDNA fragments used as a pool of templates (Lee et al., *J. Molec. Catalysis* 26:119-129 (2003)); Degenerate Oligonucleotide Gene Shuffling (DOGS), in which degenerate primers are used to control recombination between molecules; (Bergquist and Gibbs, *Methods Mol.Biol* 352:191-204 (2007); Bergquist et al., *Biomol.Eng* 22:63-72 (2005); Gibbs et al., *Gene* 271:13-20 (2001)); Incremental Truncation for the Creation of Hybrid Enzymes (ITCHY), which creates a combinatorial library with 1 base pair deletions of a gene or gene fragment of interest (Ostermeier et al., *Proc. Natl. Acad. Sci. USA* 96:3562-3567 (1999); and Ostermeier et al., *Nat. Biotechnol.* 17:1205-1209 (1999)); Thio-Incremental Truncation for the Creation of Hybrid Enzymes (THIO-ITCHY), which is similar to ITCHY except that phosphothioate dNTPs are used to generate truncations (Lutz et al., *Nucleic Acids Res* 29:E16 (2001)); SCRATCHY, which combines two methods for recombining genes, ITCHY and DNA shuffling (Lutz et al., *Proc. Natl. Acad. Sci. USA* 98:11248-11253 (2001)); Random Drift Mutagenesis (RNDM), in which mutations made via epPCR are followed by screening/selection for those retaining usable activity (Bergquist et al., *Biomol. Eng.* 22:63-72 (2005)); Sequence Saturation Mutagenesis (SeSaM), a random mutagenesis method that generates a pool of random length fragments using random incorporation of a phosphothioate nucleotide and cleavage, which is used as a template to extend in the presence of "universal" bases such as inosine, and replication of an inosine-containing complement gives random base incorporation and, consequently, mutagenesis (Wong et al., *Biotechnol. J.* 3:74-82 (2008); Wong et al., *Nucleic Acids Res.* 32:e26 (2004); and Wong et al., *Anal. Biochem.* 341:187-189 (2005)); Synthetic Shuffling, which uses overlapping oligonucleotides designed to encode "all genetic diversity in targets" and allows a very high diversity for the shuffled progeny (Ness et al., *Nat. Biotechnol.* 20:1251-1255 (2002)); Nucleotide Exchange and Excision Technology NexT, which

exploits a combination of dUTP incorporation followed by treatment with uracil DNA glycosylase and then piperidine to perform endpoint DNA fragmentation (Muller et al., *Nucleic Acids Res.* 33:e117 (2005)).

[00126] Further methods include Sequence Homology-Independent Protein Recombination (SHIPREC), in which a linker is used to facilitate fusion between two distantly related or unrelated genes, and a range of chimeras is generated between the two genes, resulting in libraries of single-crossover hybrids (Sieber et al., *Nat. Biotechnol.* 19:456-460 (2001)); Gene Site Saturation Mutagenesis™ (GSSM™), in which the starting materials include a supercoiled double stranded DNA (dsDNA) plasmid containing an insert and two primers which are degenerate at the desired site of mutations (Kretz et al., *Methods Enzymol.* 388:3-11 (2004)); Combinatorial Cassette Mutagenesis (CCM), which involves the use of short oligonucleotide cassettes to replace limited regions with a large number of possible amino acid sequence alterations (Reidhaar-Olson et al. *Methods Enzymol.* 208:564-586 (1991); and Reidhaar-Olson et al. *Science* 241:53-57 (1988)); Combinatorial Multiple Cassette Mutagenesis (CMCM), which is essentially similar to CCM and uses epPCR at high mutation rate to identify hot spots and hot regions and then extension by CMCM to cover a defined region of protein sequence space (Reetz et al., *Angew. Chem. Int. Ed Engl.* 40:3589-3591 (2001)); the Mutator Strains technique, in which conditional *ts* mutator plasmids, utilizing the *mutD5* gene, which encodes a mutant subunit of DNA polymerase III, to allow increases of 20 to 4000-X in random and natural mutation frequency during selection and block accumulation of deleterious mutations when selection is not required (Selifonova et al., *Appl. Environ. Microbiol.* 67:3645-3649 (2001)); Low et al., *J. Mol. Biol.* 260:359-3680 (1996)).

[00127] Additional exemplary methods include Look-Through Mutagenesis (LTM), which is a multidimensional mutagenesis method that assesses and optimizes combinatorial mutations of selected amino acids (Rajpal et al., *Proc. Natl. Acad. Sci. USA* 102:8466-8471 (2005)); Gene Reassembly, which is a DNA shuffling method that can be applied to multiple genes at one time or to create a large library of chimeras (multiple mutations) of a single gene (Tunable GeneReassembly™ (TGR™) Technology supplied by Verenium Corporation), *in Silico* Protein Design Automation (PDA), which is an optimization algorithm that anchors the structurally defined protein backbone possessing a particular fold, and searches sequence space for amino acid substitutions that can stabilize the fold and overall protein energetics, and generally works most effectively on proteins

with known three-dimensional structures (Hayes et al., *Proc. Natl. Acad. Sci. USA* 99:15926-15931 (2002)); and Iterative Saturation Mutagenesis (ISM), which involves using knowledge of structure/function to choose a likely site for enzyme improvement, performing saturation mutagenesis at chosen site using a mutagenesis method such as Stratagene QuikChange (Stratagene; San Diego CA), screening/selecting for desired properties, and, using improved clone(s), starting over at another site and continue repeating until a desired activity is achieved (Reetz et al., *Nat. Protoc.* 2:891-903 (2007); and Reetz et al., *Angew. Chem. Int. Ed Engl.* 45:7745-7751 (2006)).

[00128] Any of the aforementioned methods for mutagenesis can be used alone or in any combination. Additionally, any one or combination of the directed evolution methods can be used in conjunction with adaptive evolution techniques, as described herein.

[00129] It is understood that modifications which do not substantially affect the activity of the various embodiments of this invention are also provided within the definition of the invention provided herein. Accordingly, the following examples are intended to illustrate but not limit the present invention.

EXAMPLE I

Enzyme candidates for catalyzing steps A-E of Figure 1

Alcohol kinase (Figure 1, Step A)

[00130] Alcohol kinase enzymes catalyze the transfer of a phosphate group to a hydroxyl group. Kinases that catalyze transfer of a phosphate group to an alcohol group are members of the EC 2.7.1 enzyme class. The table below lists several useful kinase enzymes in the EC 2.7.1 enzyme class.

Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name
2.7.1.1	hexokinase	2.7.1.48	uridine kinase	2.7.1.94	acylglycerol kinase
2.7.1.2	glucokinase	2.7.1.49	hydroxymethylpyrimidine kinase	2.7.1.95	kanamycin kinase
2.7.1.3	ketohexokinase	2.7.1.50	hydroxyethylthiazole kinase	2.7.1.100	S-methyl-5-thioribose kinase
2.7.1.4	fructokinase	2.7.1.51	L-fuculokinase	2.7.1.101	tagatose kinase
2.7.1.5	rhamnulokinase	2.7.1.52	fucokinase	2.7.1.102	hamamelose kinase
2.7.1.6	galactokinase	2.7.1.53	L-xylulokinase	2.7.1.103	viomycin kinase
2.7.1.7	mannokinase	2.7.1.54	D-arabinokinase	2.7.1.105	6-phosphofructo-2-kinase
2.7.1.8	glucosamine kinase	2.7.1.55	allose kinase	2.7.1.106	glucose-1,6-bisphosphate synthase
2.7.1.10	phosphoglucokinase	2.7.1.56	1-phosphofructokinase	2.7.1.107	diacylglycerol kinase
2.7.1.11	6-phosphofructokinase	2.7.1.58	2-dehydro-3-deoxygalactonokinase	2.7.1.108	dolichol kinase
2.7.1.12	gluconokinase	2.7.1.59	N-acetylglucosamine kinase	2.7.1.113	deoxyguanosine kinase
2.7.1.13	dehydrogluconokinase	2.7.1.60	N-acylmannosamine kinase	2.7.1.114	AMP—thymidine kinase
2.7.1.14	sedoheptulokinase	2.7.1.61	acyl-phosphate—hexose phosphotransferase	2.7.1.118	ADP—thymidine kinase
2.7.1.15	ribokinase	2.7.1.62	phosphoramidate—hexose phosphotransferase	2.7.1.119	hygromycin-B 7"-O-kinase
2.7.1.16	ribulokinase	2.7.1.63	polyphosphate—glucose phosphotransferase	2.7.1.121	phosphoenolpyruvate—glycerone phosphotransferase
2.7.1.17	xylulokinase	2.7.1.64	inositol 3-kinase	2.7.1.122	xylitol kinase
2.7.1.18	phosphoribokinase	2.7.1.65	scyllo-inosamine 4-kinase	2.7.1.127	inositol-trisphosphate 3-kinase
2.7.1.19	phosphoribulokinase	2.7.1.66	undecaprenol kinase	2.7.1.130	tetraacyldisaccharide 4'-kinase

Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name
2.7.1.20	adenosine kinase	2.7.1.67	1-phosphatidylinositol 4-kinase	2.7.1.134	inositol-tetrakisphosphate 1-kinase
2.7.1.21	thymidine kinase	2.7.1.68	1-phosphatidylinositol-4-phosphate 5-kinase	2.7.1.136	macrolide 2'-kinase
2.7.1.22	ribosylnicotinamide kinase	2.7.1.69	protein-Np-phosphohistidine—sugar phosphotransferase	2.7.1.137	phosphatidylinositol 3-kinase
2.7.1.23	NAD+ kinase	2.7.1.70	identical to EC 2.7.1.37.	2.7.1.138	ceramide kinase
2.7.1.24	dephospho-CoA kinase	2.7.1.71	shikimate kinase	2.7.1.140	inositol-tetrakisphosphate 5-kinase
2.7.1.25	adenylyl-sulfate kinase	2.7.1.72	streptomycin 6-kinase	2.7.1.142	glycerol—3-phosphate-glucose phosphotransferase
2.7.1.26	riboflavin kinase	2.7.1.73	inosine kinase	2.7.1.143	diphosphate-purine nucleoside kinase
2.7.1.27	erythritol kinase	2.7.1.74	deoxycytidine kinase	2.7.1.144	tagatose-6-phosphate kinase
2.7.1.28	triokinase	2.7.1.76	deoxyadenosine kinase	2.7.1.145	deoxynucleoside kinase
2.7.1.29	glycerone kinase	2.7.1.77	nucleoside phosphotransferase	2.7.1.146	ADP-dependent phosphofructokinase
2.7.1.30	glycerol kinase	2.7.1.78	polynucleotide 5'-hydroxyl-kinase	2.7.1.147	ADP-dependent glucokinase
2.7.1.31	glycerate kinase	2.7.1.79	diphosphate—glycerol phosphotransferase	2.7.1.148	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase
2.7.1.32	choline kinase	2.7.1.80	diphosphate—serine phosphotransferase	2.7.1.149	1-phosphatidylinositol-5-phosphate 4-kinase
2.7.1.33	pantothenate kinase	2.7.1.81	hydroxylysine kinase	2.7.1.150	1-phosphatidylinositol-3-phosphate 5-kinase
2.7.1.34	pantetheine kinase	2.7.1.82	ethanolamine kinase	2.7.1.151	inositol-polyphosphate multikinase

Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name	Enzyme Commission Number	Enzyme Name
2.7.1.35	pyridoxal kinase	2.7.1.83	pseudouridine kinase	2.7.1.153	phosphatidylinositol-4,5-bisphosphate 3-kinase
2.7.1.36	mevalonate kinase	2.7.1.84	alkylglycerone kinase	2.7.1.154	phosphatidylinositol-4-phosphate 3-kinase
2.7.1.39	homoserine kinase	2.7.1.85	ß-glucoside kinase	2.7.1.156	adenosylcobinamide kinase
2.7.1.40	pyruvate kinase	2.7.1.86	NADH kinase	2.7.1.157	N-acetylgalactosamine kinase
2.7.1.41	glucose-1-phosphate phosphodismutase	2.7.1.87	streptomycin 3"-kinase	2.7.1.158	inositol-pentakisphosphate 2-kinase
2.7.1.42	riboflavin phosphotransferase	2.7.1.88	dihydrostreptomycin-6-phosphate 3'a-kinase	2.7.1.159	inositol-1,3,4-trisphosphate 5/6-kinase
2.7.1.43	glucuronokinase	2.7.1.89	thiamine kinase	2.7.1.160	2'-phosphotransferase
2.7.1.44	galacturonokinase	2.7.1.90	diphosphate—fructose-6-phosphate 1-phosphotransferase	2.7.1.161	CTP-dependent riboflavin kinase
2.7.1.45	2-dehydro-3-deoxygluconokinase	2.7.1.91	sphinganine kinase	2.7.1.162	N-acetylhexosamine 1-kinase
2.7.1.46	L-arabinokinase	2.7.1.92	5-dehydro-2-deoxygluconokinase	2.7.1.163	hygromycin B 4-O-kinase
2.7.1.47	D-ribulokinase	2.7.1.93	alkylglycerol kinase	2.7.1.164	O-phosphoseryl-tRNASec kinase

[00131] Mevalonate kinase (EC 2.7.1.36) phosphorylates the terminal hydroxyl group of mevalonate. Gene candidates for this step include erg12 from *S. cerevisiae*, *mvk* from *Methanocaldococcus jannaschi, MVK* from *Homo sapeins*, and *mvk* from *Arabidopsis thaliana col*. Additional mevalonate kinase candidates include the feedback-resistant mevalonate kinase from the archeon *Methanosarcina mazei* (Primak et al, *AEM*, in press (2011)) and the Mvk protein from *Streptococcus pneumoniae* (Andreassi et al, Protein Sci, 16:983-9 (2007)). Mvk proteins from *S. cerevisiae*, *S. pneumoniae* and *M. mazei* were heterologously expressed and characterized in *E. coli* (Primak et al, *supra*). The *S. pneumoniae* mevalonate kinase was active on several alternate substrates including cylopropylmevalonate, vinylmevalonate and ethynylmevalonate (Kudoh et al, *Bioorg Med Chem* 18:1124-34 (2010)), and a subsequent study determined that the ligand binding site is selective for compact, electron-rich C(3)-substituents (Lefurgy et al, *J Biol Chem* 285:20654-63 (2010)).

Protein	GenBank ID	GI Number	Organism
erg12	CAA39359.1	3684	*Sachharomyces cerevisiae*
mvk	Q58487.1	2497517	*Methanocaldococcus jannaschii*
mvk	AAH16140.1	16359371	*Homo sapiens*
mvk	NP_851084.1	30690651	*Arabidopsis thaliana*
mvk	NP_633786.1	21227864	*Methanosarcina mazei*
mvk	NP_357932.1	15902382	*Streptococcus pneumoniae*

[00132] Glycerol kinase also phosphorylates the terminal hydroxyl group in glycerol to form glycerol-3-phosphate. This reaction occurs in several species, including *Escherichia coli, Saccharomyces cerevisiae,* and *Thermotoga maritima*. The *E. coli* glycerol kinase has been shown to accept alternate substrates such as dihydroxyacetone and glyceraldehyde (Hayashi et al., *J Biol.Chem.* 242:1030-1035 (1967)). *T. maritime* has two glycerol kinases (Nelson et al., *Nature* 399:323-329 (1999)). Glycerol kinases have been shown to have a wide range of substrate specificity. Crans and Whiteside studied glycerol kinases from four different organisms (*Escherichia coli, S. cerevisiae, Bacillus stearothermophilus,* and *Candida mycoderma*) (Crans et al., *J.Am.Chem.Soc.* 107:7008-7018 (2010); Nelson et al., *supra*, (1999)). They studied 66 different analogs of glycerol and concluded that the enzyme could accept a range of substituents in place of one terminal hydroxyl group and that the hydrogen atom at C2 could be replaced by a methyl group. Interestingly, the kinetic constants of the enzyme from all four organisms were very similar.

Protein	GenBank ID	GI Number	Organism

Protein	GenBank ID	GI Number	Organism
glpK	AP_003883.1	89110103	*Escherichia coli K12*
glpK1	NP_228760.1	15642775	*Thermotoga maritime MSB8*
glpK2	NP_229230.1	15642775	*Thermotoga maritime MSB8*
Gut1	NP_011831.1	82795252	*Saccharomyces cerevisiae*

[00133] Homoserine kinase is another possible candidate. This enzyme is also present in a number of organisms including *E. coli, Streptomyces sp,* and *S. cerevisiae.* Homoserine kinase from *E. coli* has been shown to have activity on numerous substrates, including, L-2-amino,1,4- butanediol, aspartate semialdehyde, and 2-amino-5-hydroxyvalerate (Huo et al., *Biochemistry* 35:16180-16185 (1996); Huo et al., *Arch.Biochem.Biophys.* 330:373-379 (1996)). This enzyme can act on substrates where the carboxyl group at the alpha position has been replaced by an ester or by a hydroxymethyl group. The gene candidates are:

Protein	GenBank ID	GI Number	Organism
thrB	BAB96580.2	85674277	*Escherichia coli K12*
SACT1DRAFT_4809	ZP_06280784.1	282871792	*Streptomyces sp. ACT-1*
Thr1	AAA35154.1	172978	*Saccharomyces serevisiae*

Phosphate lyase (Figure 1, Step B)

[00134] Phosphate lyase enzymes catalyze the conversion of alkyl phosphates to alkenes. Carbon-oxygen lyases that operate on phosphates are found in the EC 4.2.3 enzyme class. The table below lists several relevant enzymes in EC class 4.2.3.

Enzyme Commission Number	Enzyme Name
4.2.3.5	Chorismate synthase
4.2.3.15	Myrcene synthase
4.2.3.26	Linalool synthase
4.2.3.27	Isoprene synthase
4.2.3.36	Terpentriene sythase
4.2.3.46	(E, E)-alpha-Farnesene synthase
4.2.3.47	Beta-Farnesene synthase
4.2.3.49	Nerolidol synthase

[00135] Chorismate synthase (EC 4.2.3.5) participates in the shikimate pathway, catalyzing the dephosphorylation of 5-enolpyruvylshikimate-3-phosphate to chorismate. The enzyme requires reduced flavin mononucleotide (FMN) as a cofactor, although the net reaction of the enzyme does not involve a redox change. In contrast to the enzyme found

in plants and bacteria, the chorismate synthase in fungi is also able to reduce FMN at the expense of NADPH (Macheroux et al., *Planta* 207:325-334 (1999)). Representative monofunctional enzymes are encoded by *aroC* of *E. coli* (White et al., *Biochem. J.* 251:313-322 (1988)) and *Streptococcus pneumoniae* (Maclean and Ali, *Structure* 11:1499-1511 (2003)). Bifunctional fungal enzymes are found in *Neurospora crassa* (Kitzing et al., *J. Biol. Chem.* 276:42658-42666 (2001)) and *Saccharomyces cerevisiae* (Jones et al., *Mol. Microbiol.* 5:2143-2152 (1991)).

Gene	GenBank Accession No.	GI No.	Organism
aroC	NP_416832.1	16130264	*Escherichia coli*
aroC	ACH47980.1	197205483	*Streptococcus pneumoniae*
U25818.1:19..1317	AAC49056.1	976375	*Neurospora crassa*
ARO2	CAA42745.1	3387	*Saccharomyces cerevisiae*

[00136] Isoprene synthase naturally catalyzes the conversion of dimethylallyl diphosphate to isoprene, but can also catalyze the synthesis of 1,3-butadiene from 2-butenyl-4-diphosphate. Isoprene synthases can be found in several organisms including *Populus alba* (Sasaki et al., FEBS Letters, 2005, 579 (11), 2514-2518), *Pueraria montana* (Lindberg et al., *Metabolic Eng*, 12(1):70-79 (2010); Sharkey et al., *Plant Physiol.*, 137(2):700-712 (2005)), and *Populus tremula* x *Populus alba*, also called *Populus canescens* (Miller et al., Planta, 2001, 213 (3), 483-487). The crystal structure of the *Populus canescens* isoprene synthase was determined (Koksal et al, J Mol Biol 402:363-373 (2010)). Additional isoprene synthase enzymes are described in (Chotani et al., WO/2010/031079, Systems Using Cell Culture for Production of Isoprene; Cervin et al., US Patent Application 20100003716, Isoprene Synthase Variants for Improved Microbial Production of Isoprene).

Protein	GenBank ID	GI Number	Organism
ispS	BAD98243.1	63108310	*Populus alba*
ispS	AAQ84170.1	35187004	*Pueraria montana*
ispS	CAC35696.1	13539551	*Populus tremula x Populus alba*

[00137] Myrcene synthase enzymes catalyze the dephosphorylation of geranyl diphosphate to beta-myrcene (EC 4.2.3.15). Exemplary myrcene synthases are encoded by MST2 of *Solanum lycopersicum* (van Schie et al, Plant Mol Biol 64:D473-79 (2007)), TPS-Myr of *Picea abies* (Martin et al, Plant Physiol 135:1908-27 (2004)) g-myr of *Abies*

grandis (Bohlmann et al, J Biol Chem 272:21784-92 (1997)) and TPS10 of *Arabidopsis thaliana* (Bohlmann et al, Arch Biochem Biophys 375:261-9 (2000)). These enzymes were heterologously expressed in *E. coli*.

Protein	GenBank ID	GI Number	Organism
MST2	ACN58229.1	224579303	*Solanum lycopersicum*
TPS-Myr	AAS47690.2	77546864	*Picea abies*
G-myr	O24474.1	17367921	*Abies grandis*
TPS10	EC07543.1	330252449	*Arabidopsis thaliana*

[00138] Farnesyl diphosphate is converted to alpha-farnesene and beta-farnesene by alpha-farnesene synthase and beta-farnesene synthase, respectively. Exemplary alpha-farnesene synthase enzymes include *TPS03* and *TPS02* of Arabidopsis thaliana (Faldt et al, *Planta* 216:745-51 (2003); Huang et al, *Plant Physiol* 153:1293-310 (2010)), *afs* of *Cucumis sativus* (Mercke et al, Plant Physiol 135:2012-14 (2004), *eafar* of *Malus x domestica* (Green et al, Phytochem 68:176-88 (2007)) and TPS-Far of Picea abies (Martin, *supra*). An exemplary beta-farnesene synthase enzyme is encoded by *TPS1* of *Zea mays* (Schnee et al, Plant Physiol 130:2049-60 (2002)).

Protein	GenBank ID	GI Number	Organism
TPS03	A4FVP2.1	205829248	*Arabidopsis thaliana*
TPS02	P0CJ43.1	317411866	*Arabidopsis thaliana*
TPS-Far	AAS47697.1	44804601	*Picea abies*
afs	AAU05951.1	51537953	*Cucumis sativus*
eafar	Q84LB2.2	75241161	*Malus x domestica*
TPS1	Q84ZW8.1	75149279	*Zea mays*

Diphosphokinase (Figure 1, Step C)

[00139] Diphosphokinase enzymes catalyze the transfer of a diphosphate group to an alcohol group. The enzymes described below naturally possess such activity. Kinases that catalyze transfer of a diphosphate group are members of the EC 2.7.6 enzyme class. The table below lists several useful kinase enzymes in the EC 2.7.6 enzyme class.

Enzyme Commission No.	Enzyme Name
2.7.6.1	ribose-phosphate diphosphokinase
2.7.6.2	thiamine diphosphokinase
2.7.6.3	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase
2.7.6.4	nucleotide diphosphokinase

Enzyme Commission No.	Enzyme Name
2.7.6.5	GTP diphosphokinase

[00140] Of particular interest are ribose-phosphate diphosphokinase enzymes, which have been identified in *Escherichia coli* (Hove-Jenson et al., *J Biol Chem*, 1986, 261(15);6765-71) and *Mycoplasma pneumoniae* M129 (McElwain et al, *International Journal of Systematic Bacteriology*, 1988, 38:417-423) as well as thiamine diphosphokinase enzymes. Exemplary thiamine diphosphokinase enzymes are found in *Arabidopsis thaliana* (Ajjawi, *Plant Mol Biol*, 2007, 65(1-2);151-62).

Protein	GenBank ID	GI Number	Organism
prs	NP_415725.1	*16129170*	*Escherichia coli*
prsA	NP_109761.1	*13507812*	*Mycoplasma pneumoniae M129*
TPK1	BAH19964.1	*222424006*	*Arabidopsis thaliana col*
TPK2	BAH57065.1	*227204427*	*Arabidopsis thaliana col*

Alkyl phosphate kinase (Figure 1, Step D)

[00141] Alkyl phosphate kinase enzymes catalyze the transfer of a phosphate group to the phosphate group of an alkyl phosphate. The enzymes described below naturally possess such activity or can be engineered to exhibit this activity. Kinases that catalyze transfer of a phosphate group to another phosphate group are members of the EC 2.7.4 enzyme class. The table below lists several useful kinase enzymes in the EC 2.7.4 enzyme class.

Enzyme Commission No.	Enzyme Name
2.7.4.1	polyphosphate kinase
2.7.4.2	phosphomevalonate kinase
2.7.4.3	adenylate kinase
2.7.4.4	nucleoside-phosphate kinase
2.7.4.6	nucleoside-diphosphate kinase
2.7.4.7	phosphomethylpyrimidine kinase
2.7.4.8	guanylate kinase
2.7.4.9	dTMP kinase
2.7.4.10	nucleoside-triphosphate—adenylate kinase
2.7.4.11	(deoxy)adenylate kinase
2.7.4.12	T2-induced deoxynucleotide kinase
2.7.4.13	(deoxy)nucleoside-phosphate kinase
2.7.4.14	cytidylate kinase
2.7.4.15	thiamine-diphosphate kinase

Enzyme Commission No.	Enzyme Name
2.7.4.16	thiamine-phosphate kinase
2.7.4.17	3-phosphoglyceroyl-phosphate—polyphosphate phosphotransferase
2.7.4.18	farnesyl-diphosphate kinase
2.7.4.19	5-methyldeoxycytidine-5'-phosphate kinase
2.7.4.20	dolichyl-diphosphate—polyphosphate phosphotransferase
2.7.4.21	inositol-hexakisphosphate kinase
2.7.4.22	UMP kinase
2.7.4.23	ribose 1,5-bisphosphate phosphokinase
2.7.4.24	diphosphoinositol-pentakisphosphate kinase
2.7.4.-	Farnesyl monophosphate kinase
2.7.4.-	Geranyl-geranyl monophosphate kinase
2.7.4.-	Phytyl-phosphate kinase

[00142] Phosphomevalonate kinase enzymes are of particular interest. Phosphomevalonate kinase (EC 2.7.4.2) catalyzes the phosphorylation of phosphomevalonate. This enzyme is encoded by *erg8* in *Saccharomyces cerevisiae* (Tsay et al., *Mol.Cell Biol.* 11:620-631 (1991)) and *mvaK2* in *Streptococcus pneumoniae, Staphylococcus aureus* and *Enterococcus faecalis* (Doun et al., *Protein Sci.* 14:1134-1139 (2005); Wilding et al., *J Bacteriol.* 182:4319-4327 (2000)). The *Streptococcus pneumoniae* and *Enterococcus faecalis* enzymes were cloned and characterized in *E. coli* (Pilloff et al., *J Biol.Chem.* 278:4510-4515 (2003); Doun et al., *Protein Sci.* 14:1134-1139 (2005)). The *S. pneumoniae* phosphomevalonate kinase was active on several alternate substrates including cylopropylmevalonate phosphate, vinylmevalonate phosphate and ethynylmevalonate phosphate (Kudoh et al, *Bioorg Med Chem* 18:1124-34 (2010)).

Protein	GenBank ID	GI Number	Organism
Erg8	AAA34596.1	171479	*Saccharomyces cerevisiae*
mvaK2	AAG02426.1	9937366	*Staphylococcus aureus*
mvaK2	AAG02457.1	9937409	*Streptococcus pneumoniae*
mvaK2	AAG02442.1	9937388	*Enterococcus faecalis*

[00143] Farnesyl monophosphate kinase enzymes catalyze the CTP dependent phosphorylation of farnesyl monophosphate to farnesyl diphosphate. Similarly, geranylgeranyl phosphate kinase catalyzes CTP dependent phosphorylation. Enzymes with these activities were identified in the microsomal fraction of cultured *Nicotiana tabacum*

(Thai et al, PNAS 96:13080-5 (1999)). However, the associated genes have not been identified to date.

Diphosphate lyase (Figure 1, Step E)

[00144] Diphosphate lyase enzymes catalyze the conversion of alkyl diphosphates to alkenes. Carbon-oxygen lyases that operate on phosphates are found in the EC 4.2.3 enzyme class. The table below lists several useful enzymes in EC class 4.2.3. Exemplary enzyme candidates were described above (see phosphate lyase section).

Enzyme Commission No.	Enzyme Name
4.2.3.5	Chorismate synthase
4.2.3.15	Myrcene synthase
4.2.3.27	Isoprene synthase
4.2.3.36	Terpentriene sythase
4.2.3.46	(E, E)-alpha-Farnesene synthase
4.2.3.47	Beta-Farnesene synthase

EXAMPLE II

Preparation of an Isobutylene Producing Microbial Organism

[00145] This example describes the generation of a microbial organism capable of producing isobutylene from isobutanol, in an organism engineered to have an isobutylene pathway.

[00146] An isobutanol-overproducing strain of *Escherichia coli* is used as a target organism to engineer an isobutylene-producing pathway. Pathways for efficiently converting central metabolic intermediates to isobutanol are known in the art (for example: US 8017375; PCT/US2006/041602; PCT/US2008/053514; PCT/US2006/041602; Dickinson et al., JBC 273:25751-56 (1998)) and isobutanol overproducing *E. coli* strains have been developed (for example, Atsumi et al, Appl Microbiol Biotech 85:651-57 (2010)).

[00147] To generate an *E. coli* strain engineered to produce isobutylene from isobutanol, nucleic acids encoding the enzymes utilized in the pathway of Figure 1 are expressed in *E. coli* using well known molecular biology techniques (see, for example, Sambrook, *supra*, 2001; Ausubel *supra*, 1999; Roberts et al., *supra*, 1989). In particular, the *mvk* (NP_357932.1), *mvaK2* (AAG02457.1) and, *ispS* (CAC35696.1) genes encoding

alkyl phosphate kinase, alkyl diphosphate kinase and isobutylene synthetase, respectively, are cloned into the pZE13 vector (Expressys, Ruelzheim, Germany), under the control of the PA1/lacO promoter. This plasmid is then transformed into a host strain containing $lacI^Q$, which allows inducible expression by addition of isopropyl-beta-D-1-thiogalactopyranoside (IPTG).

[00148] The resulting genetically engineered organism is cultured in glucose containing medium following procedures well known in the art (see, for example, Sambrook et al., *supra*, 2001). The expression of isobutylene pathway genes is corroborated using methods well known in the art for determining polypeptide expression or enzymatic activity, including for example, Northern blots, PCR amplification of mRNA and immunoblotting. Enzymatic activities of the expressed enzymes are confirmed using assays specific for the individually activities. The ability of the engineered *E. coli* strain to produce isobutylene is confirmed using HPLC, gas chromatography-mass spectrometry (GCMS) or liquid chromatography-mass spectrometry (LCMS).

[00149] Microbial strains engineered to have a functional isobutylene synthesis pathway are further augmented by optimization for efficient utilization of the pathway. Briefly, the engineered strain is assessed to determine whether any of the exogenous genes are expressed at a rate limiting level. Expression is increased for any enzymes expressed at low levels that can limit the flux through the pathway by, for example, introduction of additional gene copy numbers. Strategies are also applied to improve production of isobutylene precursor isobutanoyl-phosphate, such as mutagenesis, cloning and/or deletion of native genes involved in byproduct formation.

[00150] To generate better producers, metabolic modeling is utilized to optimize growth conditions. Modeling is also used to design gene knockouts that additionally optimize utilization of the pathway (see, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and in U.S. Patent No. 7,127,379). Modeling analysis allows reliable predictions of the effects on cell growth of shifting the metabolism towards more efficient production of isobutylene. One modeling method is the bilevel optimization approach, OptKnock (Burgard et al., *Biotechnol. Bioengineer.* 84:647-657 (2003)), which is applied to select gene knockouts that collectively result in better production of isobutylene. Adaptive evolution also can be used to generate better producers of, for example, the isobutanoyl-phosphate intermediate

or the isobutylene product. Adaptive evolution is performed to improve both growth and production characteristics (Fong and Palsson, *Nat. Genet.* 36:1056-1058 (2004); Alper et al., *Science* 314:1565-1568 (2006)). Based on the results, subsequent rounds of modeling, genetic engineering and adaptive evolution can be applied to the isobutylene producer to further increase production.

[00151] For large-scale production of isobutylene, the above isobutylene pathway-containing organism is cultured in a fermenter using a medium known in the art to support growth of the organism under anaerobic conditions. Fermentations are performed in either a batch, fed-batch or continuous manner. Anaerobic conditions are maintained by first sparging the medium with nitrogen and then sealing culture vessel (e.g., flasks can be sealed with a septum and crimp-cap). Microaerobic conditions also can be utilized by providing a small hole for limited aeration. The pH of the medium is maintained at a pH of 7 by addition of an acid, such as H2SO4. The growth rate is determined by measuring optical density using a spectrophotometer (600 nm), and the glucose uptake rate by monitoring carbon source depletion over time. Byproducts such as undesirable alcohols, organic acids, and residual glucose can be quantified by HPLC (Shimadzu) with an HPX-087 column (BioRad), using a refractive index detector for glucose and alcohols, and a UV detector for organic acids, Lin et al., *Biotechnol. Bioeng.*, 775-779 (2005).

EXAMPLE III

Pathways for the formation of butadiene precursor 3-buten-1-ol (but-3-en-1-ol) from pyruvate and acetaldehyde.

[00152] This example describes pathways for converting pyruvate and acetaldehyde to 3-buten-1-ol, and further to butadiene. The conversion of pyruvate and acetaldehyde to 3-buten-1-ol is accomplished in four enzymatic steps. Pyruvate and acetaldehyde are first condensed to 4-hydroxy-2-oxovalerate by 4-hydroxy-2-ketovalerate aldolase. The 4-hydroxy-2-oxovalerate product is subsequently dehydrated to 2-oxopentenoate. Decarboxylation of 2-oxopentenoate yields 3-buten-1-al, which is further reduced to 3-buten-1-ol by an alcohol dehydrogenase.

[00153] Enzymes and gene candidates for catalyzing but-3-en-1-ol pathway reactions are described in further detail below.

[00154] The condensation of pyruvate and acetaldehyde to 4-hydroxy-2-oxovalerate is catalyzed by 4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39). This enzyme participates in pathways for the degradation of phenols, cresols and catechols. The *E. coli* enzyme, encoded by *mhpE*, is highly specific for acetaldehyde as an acceptor but accepts the alternate substrates 2-ketobutyrate or phenylpyruvate as donors (Pollard et al., *Appl Environ Microbiol* 64:4093-4094 (1998)). Similar enzymes are encoded by the *cmtG* and *todH* genes of *Pseudomonas putida* (Lau et al., *Gene* 146:7-13 (1994); Eaton, *J Bacteriol.* 178:1351-1362 (1996)). In *Pseudomonas* CF600, this enzyme is part of a bifunctional aldolase-dehydrogenase heterodimer encoded by *dmpFG (Manjasetty et al., Acta Crystallogr.D.Biol Crystallogr. 57:582-585 (2001))*. The dehydrogenase functionality interconverts acetaldehyde and acetyl-CoA, providing the advantage of reduced cellular concentrations of acetaldehyde, toxic to some cells.

Gene	GenBank ID	GI Number	Organism
mhpE	AAC73455.1	1786548	*Escherichia coli*
cmtG	AAB62295.1	1263190	*Pseudomonas putida*
todH	AAA61944.1	485740	*Pseudomonas putida*
dmpG	CAA43227.1	45684	*Pseudomonas sp. CF600*
dmpF	CAA43226.1	45683	*Pseudomonas sp. CF600*

[00155] Dehydration of 4-hydroxy-2-oxovalerate to 2-oxopentenoate is catalyzed by 4-hydroxy-2-oxovalerate hydratase (EC 4.2.1.80). This enzyme participates in aromatic degradation pathways and is typically co-transcribed with a gene encoding an enzyme with 4-hydroxy-2-oxovalerate aldolase activity. Exemplary gene products are encoded by *mhpD* of *E. coli* (Ferrandez et al., *J Bacteriol.* 179:2573-2581 (1997); Pollard et al., *Eur J Biochem.* 251:98-106 (1998)), *todG* and *cmtF* of *Pseudomonas putida* (Lau et al., *Gene* 146:7-13 (1994); Eaton, *J Bacteriol.* 178:1351-1362 (1996)), *cnbE* of *Comamonas* sp. CNB-1 (Ma et al., *Appl Environ Microbiol* 73:4477-4483 (2007)) and mhpD of *Burkholderia xenovorans* (Wang et al., *FEBS J* 272:966-974 (2005)). A closely related enzyme, 2-oxohepta-4-ene-1,7-dioate hydratase, participates in 4-hydroxyphenylacetic acid degradation, where it converts 2-oxo-hept-4-ene-1,7-dioate (OHED) to 2-oxo-4-hydroxy-hepta-1,7-dioate using magnesium as a cofactor (Burks et al., *J.Am.Chem.Soc.* 120: (1998)). OHED hydratase enzyme candidates have been identified and characterized in *E. coli C* (Roper et al., *Gene* 156:47-51 (1995); Izumi et al., *J Mol.Biol.* 370:899-911 (2007)) and *E. coli W* (Prieto et al., *J Bacteriol.* 178:111-120 (1996)). Sequence comparison reveals homologs in a wide range of bacteria, plants and animals. Enzymes

with highly similar sequences are contained in *Klebsiella pneumonia* (91% identity, eval = 2e-138) and *Salmonella enterica* (91% identity, eval = 4e-138), among others.

Protein	GenBank Accession No.	GI No.	Organism
mhpD	AAC73453.2	87081722	*Escherichia coli*
cmtF	AAB62293.1	1263188	*Pseudomonas putida*
todG	AAA61942.1	485738	*Pseudomonas putida*
cnbE	YP_001967714.1	190572008	*Comamonas* sp. CNB-1
mhpD	Q13VU0	123358582	*Burkholderia xenovorans*
hpcG	CAA57202.1	556840	*Escherichia coli C*
hpaH	CAA86044.1	757830	*Escherichia coli W*
hpaH	ABR80130.1	150958100	*Klebsiella pneumoniae*
Sari_01896	ABX21779.1	160865156	*Salmonella enterica*

[00156] Decarboxylation of 4-hydroxy-2-oxovalerate is catalyzed by a keto-acid decarboxylase. Suitable enzyme candidates include pyruvate decarboxylase (EC 4.1.1.1), benzoylformate decarboxylase (EC 4.1.1.7), alpha-ketoglutarate decarboxylase and branched-chain alpha-ketoacid decarboxylase. Pyruvate decarboxylase (PDC), also termed keto-acid decarboxylase, is a key enzyme in alcoholic fermentation, catalyzing the decarboxylation of pyruvate to acetaldehyde. The enzyme from *Saccharomyces cerevisiae* has a broad substrate range for aliphatic 2-keto acids including 2-ketobutyrate, 2-ketovalerate, 3-hydroxypyruvate and 2-phenylpyruvate (22). This enzyme has been extensively studied, engineered for altered activity, and functionally expressed in *E. coli* (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001); Li et al., *Biochemistry.* 38:10004-10012 (1999); ter Schure et al., *Appl.Environ.Microbiol.* 64:1303-1307 (1998)). The PDC from *Zymomonas mobilus,* encoded by *pdc*, also has a broad substrate range and has been a subject of directed engineering studies to alter the affinity for different substrates (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)). The crystal structure of this enzyme is available (Killenberg-Jabs et al., *Eur.J.Biochem.* 268:1698-1704 (2001)). Other well-characterized PDC candidates include the enzymes from *Acetobacter pasteurians* (Chandra et al., 176:443-451 (2001)) and *Kluyveromyces lactis* (Krieger et al., 269:3256-3263 (2002)).

Protein	GenBank ID	GI Number	Organism
pdc	P06672.1	118391	*Zymomonas mobilis*
pdc1	P06169	30923172	*Saccharomyces cerevisiae*
pdc	Q8L388	20385191	*Acetobacter pasteurians*
pdc1	Q12629	52788279	*Kluyveromyces lactis*

[00157] Like PDC, benzoylformate decarboxylase (EC 4.1.1.7) has a broad substrate range and has been the target of enzyme engineering studies. The enzyme from *Pseudomonas putida* has been extensively studied and crystal structures of this enzyme are available (Polovnikova et al., 42:1820-1830 (2003); Hasson et al., 37:9918-9930 (1998)). Site-directed mutagenesis of two residues in the active site of the *Pseudomonas putida* enzyme altered the affinity (Km) of naturally and non-naturally occurring substrates (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)). The properties of this enzyme have been further modified by directed engineering (Lingen et al., *Chembiochem.* 4:721-726 (2003); Lingen et al., *Protein Eng* 15:585-593 (2002)). The enzyme from *Pseudomonas aeruginosa,* encoded by *mdlC,* has also been characterized experimentally (Barrowman et al., 34:57-60 (1986)). Additional gene candidates from *Pseudomonas stutzeri*, *Pseudomonas fluorescens* and other organisms can be inferred by sequence homology or identified using a growth selection system developed in *Pseudomonas putida* (Henning et al., *Appl.Environ.Microbiol.* 72:7510-7517 (2006)).

Protein	GenBank ID	GI Number	Organism
mdlC	P20906.2	3915757	*Pseudomonas putida*
mdlC	Q9HUR2.1	81539678	*Pseudomonas aeruginosa*
dpgB	ABN80423.1	126202187	*Pseudomonas stutzeri*
ilvB-1	YP_260581.1	70730840	*Pseudomonas fluorescens*

[00158] A third enzyme capable of decarboxylating 2-oxoacids is alpha-ketoglutarate decarboxylase (KGD). The substrate range of this class of enzymes has not been studied to date. An exemplarly KDC is encoded by *kad* in *Mycobacterium tuberculosis* (Tian et al., *PNAS* 102:10670-10675 (2005)). KDC enzyme activity has also been detected in several species of rhizobia including *Bradyrhizobium japonicum* and *Mesorhizobium loti* (Green et al., *J Bacteriol* 182:2838-2844 (2000)). Although the KDC-encoding gene(s) have not been isolated in these organisms, the genome sequences are available and several genes in each genome are annotated as putative KDCs. A KDC from *Euglena gracilis* has also been characterized but the gene associated with this activity has not been identified to date (Shigeoka et al., *Arch.Biochem.Biophys.* 288:22-28 (1991)). The first twenty amino acids starting from the N-terminus were sequenced MTYKAPVKDVKFLLDKVFKV (SEQ ID NO.) (Shigeoka and Nakano, *Arch.Biochem.Biophys.* 288:22-28 (1991)). The gene could

be identified by testing candidate genes containing this N-terminal sequence for KDC activity.

Protein	GenBank ID	GI Number	Organism
kgd	O50463.4	160395583	*Mycobacterium tuberculosis*
kgd	NP_767092.1	27375563	*Bradyrhizobium japonicum USDA110*
kgd	NP_105204.1	13473636	*Mesorhizobium loti*

[00159] A fourth candidate enzyme for catalyzing this reaction is branched chain alpha-ketoacid decarboxylase (BCKA). This class of enzyme has been shown to act on a variety of compounds varying in chain length from 3 to 6 carbons (Oku et al., *J Biol Chem.* 263:18386-18396 (1988); Smit et al., *Appl Environ Microbiol* 71:303-311 (2005)). The enzyme in *Lactococcus lactis* has been characterized on a variety of branched and linear substrates including 2-oxobutanoate, 2-oxohexanoate, 2-oxopentanoate, 3-methyl-2-oxobutanoate, 4-methyl-2-oxobutanoate and isocaproate (Smit et al., *Appl Environ Microbiol* 71:303-311 (2005)). The enzyme has been structurally characterized (Berg et al., *Science.* 318:1782-1786 (2007)). Sequence alignments between the *Lactococcus lactis* enzyme and the pyruvate decarboxylase of *Zymomonas mobilus* indicate that the catalytic and substrate recognition residues are nearly identical (Siegert et al., *Protein Eng Des Sel* 18:345-357 (2005)), so this enzyme would be a promising candidate for directed engineering. Decarboxylation of alpha-ketoglutarate by a BCKA was detected in *Bacillus subtilis*; however, this activity was low (5%) relative to activity on other branched-chain substrates (Oku and Kaneda, *J Biol Chem.* 263:18386-18396 (1988)) and the gene encoding this enzyme has not been identified to date. Additional BCKA gene candidates can be identified by homology to the *Lactococcus lactis* protein sequence. Many of the high-scoring BLASTp hits to this enzyme are annotated as indolepyruvate decarboxylases (EC 4.1.1.74). Indolepyruvate decarboxylase (IPDA) is an enzyme that catalyzes the decarboxylation of indolepyruvate to indoleacetaldehyde in plants and plant bacteria. Recombinant branched chain alpha-keto acid decarboxylase enzymes derived from the E1 subunits of the mitochondrial branched-chain keto acid dehydrogenase complex from *Homo sapiens* and *Bos taurus* have been cloned and functionally expressed in *E. coli* (Davie et al., *J.Biol.Chem.* 267:16601-16606 (1992); Wynn et al., *J.Biol.Chem.* 267:12400-12403 (1992); Wynn et al., *J.Biol.Chem.* 267:1881-1887 (1992)). In these studies, the authors found that co-expression of chaperonins GroEL and GroES enhanced the specific activity of the decarboxylase by 500-fold (Wynn et al., *J.Biol.Chem.*

267:12400-12403 (1992)). These enzymes are composed of two alpha and two beta subunits.

Protein	GenBank ID	GI Number	Organism
kdcA	AAS49166.1	44921617	*Lactococcus lactis*
BCKDHB	NP_898871.1	34101272	*Homo sapiens*
BCKDHA	NP_000700.1	11386135	*Homo sapiens*
BCKDHB	P21839	115502434	*Bos taurus*
BCKDHA	P11178	129030	*Bos taurus*

[00160] Reduction of 3-buten-1-al to 3-buten-1-ol is catalyzed by an aldehyde reductase or alcohol dehydrogenase. Genes encoding enzymes that catalyze the reduction of an aldehyde to alcohol (i.e., alcohol dehydrogenase or equivalently aldehyde reductase) include *alrA* encoding a medium-chain alcohol dehydrogenase for C2-C14 (Tani et al., *Appl.Environ.Microbiol.* 66:5231-5235 (2000)), *yqhD* and *fucO* from *E. coli* (Sulzenbacher et al., 342:489-502 (2004)), and *bdh* I and *bdh* II from *C. acetobutylicum* which converts butyryaldehyde into butanol (Walter et al., 174:7149-7158 (1992)). YqhD catalyzes the reduction of a wide range of aldehydes using NADPH as the cofactor, with a preference for chain lengths longer than C(3) (Sulzenbacher et al., 342:489-502 (2004);Perez et al., *J Biol.Chem.* 283:7346-7353 (2008)). The *adhA* gene product from *Zymomonas mobilisE* has been demonstrated to have activity on a number of aldehydes including formaldehyde, acetaldehyde, propionaldehyde, butyraldehyde, and acrolein (Kinoshita et al., *Appl Microbiol Biotechnol* 22:249-254 (1985)). Additional aldehyde reductase candidates are encoded by *bdh* in *C. saccharoperbutylacetonicum* and *Cbei_1722, Cbei_2181* and *Cbei_2421* in *C. Beijerinckii.* Additional aldehyde reductase gene candidates in *Saccharomyces cerevisiae* include the aldehyde reductases GRE3, ALD2-6 and HFD1, glyoxylate reductases GOR1 and YPL113C and glycerol dehydrogenase GCY1 (WO 2011/022651A1; Atsumi et al., *Nature* 451:86-89 (2008)). The enzyme candidates described previously for catalyzing the reduction of methylglyoxal to acetol or lactaldehyde are also suitable lactaldehyde reductase enzyme candidates.

Protein	GENBANK ID	GI NUMBER	ORGANISM
alrA	BAB12273.1	9967138	*Acinetobacter* sp. strain M-1
ADH2	NP_014032.1	6323961	*Saccharomyces cerevisiae*
yqhD	NP_417484.1	16130909	*Escherichia coli*
fucO	NP_417279.1	16130706	*Escherichia coli*
bdh I	NP_349892.1	15896543	*Clostridium acetobutylicum*
bdh II	NP_349891.1	15896542	*Clostridium acetobutylicum*

adhA	YP_162971.1	56552132	*Zymomonas mobilis*
bdh	BAF45463.1	124221917	*Clostridium saccharoperbutylacetonicum*
Cbei_1722	YP_001308850	150016596	*Clostridium beijerinckii*
Cbei_2181	YP_001309304	150017050	*Clostridium beijerinckii*
Cbei_2421	YP_001309535	150017281	*Clostridium beijerinckii*
GRE3	P38715.1	731691	*Saccharomyces cerevisiae*
ALD2	CAA89806.1	825575	*Saccharomyces cerevisiae*
ALD3	NP_013892.1	6323821	*Saccharomyces cerevisiae*
ALD4	NP_015019.1	6324950	*Saccharomyces cerevisiae*
ALD5	NP_010996.2	330443526	*Saccharomyces cerevisiae*
ALD6	ABX39192.1	160415767	*Saccharomyces cerevisiae*
HFD1	Q04458.1	2494079	*Saccharomyces cerevisiae*
GOR1	NP_014125.1	6324055	*Saccharomyces cerevisiae*
YPL113C	AAB68248.1	1163100	*Saccharomyces cerevisiae*
GCY1	CAA99318.1	1420317	*Saccharomyces cerevisiae*

[00161] Enzymes exhibiting 4-hydroxybutyrate dehydrogenase activity (EC 1.1.1.61) and glutarate semialdehyde reductase also fall into this category. 4-Hydroxybutyrate dehydrogenase enzymes have been characterized in *Ralstonia eutropha* (Bravo et al., *J Forens Sci*, 49:379-387 (2004)) and *Clostridium kluyveri* (Wolff et al., *Protein Expr.Purif.* 6:206-212 (1995)). Yet another gene is the alcohol dehydrogenase *adhI* from *Geobacillus thermoglucosidasius* (Jeon et al., *J Biotechnol* 135:127-133 (2008)). Glutarate semialdehyde reductase enzymes include the ATEG_00539 gene product of *Aspergillus terreus* and 4-hydroxybutyrate dehydrogenase of *Arabidopsis thaliana,* encoded by *4hbd* (WO 2010/068953A2). The *A. thaliana* enzyme was cloned and characterized in yeast (Breitkreuz et al., *J.Biol.Chem.* 278:41552-41556 (2003)).

PROTEIN	GENBANK ID	GI NUMBER	ORGANISM
4hbd	YP_726053.1	113867564	*Ralstonia eutropha* H16
4hbd	L21902.1	146348486	*Clostridium kluyveri* DSM 555
adhI	AAR91477.1	40795502	*Geobacillus thermoglucosidasius*
ATEG_00539	XP_001210625.1	115491995	*Aspergillus terreus NIH2624*
4hbd	AAK94781.1	15375068	*Arabidopsis thaliana*

[00162] Another exemplary aldehyde reductase is methylmalonate semialdehyde reductase, also known as 3-hydroxyisobutyrate dehydrogenase (EC 1.1.1.31). This enzyme participates in valine, leucine and isoleucine degradation and has been identified in bacteria, eukaryotes, and mammals. The enzyme encoded by *P84067* from *Thermus thermophilus* HB8 has been structurally characterized (Lokanath et al., *J Mol Biol*, 352:905-17 (2005)). The reversibility of the human 3-hydroxyisobutyrate dehydrogenase

was demonstrated using isotopically-labeled substrate (Manning et al., *Biochem J*, 231:481-4 (1985)). Additional genes encoding this enzyme include *3hidh* in *Homo sapiens* (Hawes et al., *Methods Enzymol*, 324:218-228 (2000)) and *Oryctolagus cuniculus* (Hawes et al., *supra*; Chowdhury et al., *Biosci.Biotechnol Biochem.* 60:2043-2047 (1996)), *mmsB* in *Pseudomonas aeruginosa* and *Pseudomonas putida*, and *dhat* in *Pseudomonas putida* (Aberhart et al., *J Chem.Soc.[Perkin 1]* 6:1404-1406 (1979); Chowdhury et al., *Biosci.Biotechnol Biochem.* 60:2043-2047 (1996); Chowdhury et al., *Biosci.Biotechnol Biochem.* 67:438-441 (2003)). Several 3-hydroxyisobutyrate dehydrogenase enzymes have been characterized in the reductive direction, including *mmsB* from *Pseudomonas aeruginosa* (Gokarn et al., US Patent 739676, (2008)) and *mmsB* from *Pseudomonas putida*.

PROTEIN	GENBANK ID	GI NUMBER	ORGANISM
P84067	P84067	75345323	*Thermus thermophilus*
3hidh	P31937.2	12643395	*Homo sapiens*
3hidh	P32185.1	416872	*Oryctolagus cuniculus*
mmsB	NP_746775.1	26991350	*Pseudomonas putida*
mmsB	P28811.1	127211	*Pseudomonas aeruginosa*
dhat	Q59477.1	2842618	*Pseudomonas putida*

Example IV

Preparation of a Butadiene Producing Microbial Organism with a But-3-en-1-ol Pathway

[00163] This example describes the generation of a microbial organism capable of producing butadiene from pyruvate via a but-3-en-1-ol intermediate, in an organism engineered to have a butadiene pathway.

[00164] *Escherichia coli* is used as a target organism to engineer a butadiene-producing pathway. *E. coli* provides a good host for generating a non-naturally occurring microorganism capable of producing butadiene. *E. coli* is amenable to genetic manipulation and is known to be capable of producing various products, including ethanol, acetic acid, formic acid, lactic acid, and succinic acid, effectively under anaerobic or microaerobic conditions.

[00165] To generate an *E. coli* strain engineered to produce the butadiene pathway precursor, but-3-en-1-ol, a functional nucleic acids encoding the enzymes utilized in the pathway described in Example III, are expressed in *E. coli* using well known molecular biology techniques (see, for example, Sambrook, *supra*, 2001; Ausubel *supra*, 1999; Roberts et al., *supra*, 1989).

[00166] In particular, an *E. coli* strain is engineered to produce but-3-en-1-ol from pyruvate via the route described in Example 3. For the first stage of pathway construction, genes encoding enzymes to transform pyruvate to but-3-en-1-ol are assembled onto a vector. The genes mhpE (AAC73455.1), mhpD (AAC73453.2), kdcA (AAS49166.1), adhA (YP_162971.1) encoding 4-hydroxy-2-oxovalerate aldolase, 4-hydroxy-2-oxovalerate dehydratase, 2-oxopentenoate decarboxylase and 3-buten-1-al reductase, respectively, are cloned into the pZE13 vector (Expressys, Ruelzheim, Germany), under the control of the PA1/lacO promoter. The genes *mvk* (NP_357932.1), *mvaK2* (AAG02457.1) and, *ispS* (CAC35696.1) encoding alkyl phosphate kinase, alkyl diphosphate kinase and butadiene synthetase, respectively, are cloned into the pZA33 vector (Expressys, Ruelzheim, Germany) under the PA1/lacO promoter. The two plasmids are transformed into *E. coli* host strain containing $lacI^Q$, which allows inducible expression by addition of isopropyl-beta-D-1-thiogalactopyranoside (IPTG).

[00167] The resulting genetically engineered organism is cultured in glucose containing medium following procedures well known in the art (see, for example, Sambrook et al., *supra*, 2001). The expression of butadiene pathway genes is corroborated using methods well known in the art for determining polypeptide expression or enzymatic activity, including for example, Northern blots, PCR amplification of mRNA and immunoblotting. Enzymatic activities of the expressed enzymes are confirmed using assays specific for the individually activities. The ability of the engineered *E. coli* strain to produce butadiene is confirmed using HPLC, gas chromatography-mass spectrometry (GCMS) or liquid chromatography-mass spectrometry (LCMS).

[00168] Microbial strains engineered to have a functional butadiene synthesis pathway are further augmented by optimization for efficient utilization of the pathway. Briefly, the engineered strain is assessed to determine whether any of the exogenous genes are expressed at a rate limiting level. Expression is increased for any enzymes expressed at low levels that can limit the flux through the pathway by, for example, introduction of additional gene copy numbers. Strategies are also applied to improve production of

butadiene precursor but-3-en-1-ol, such as mutagenesis, cloning and/or deletion of native genes involved in byproduct formation.

[00169] To generate better butadiene producers, metabolic modeling is utilized to optimize growth conditions. Modeling is also used to design gene knockouts that additionally optimize utilization of the pathway (see, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and in U.S. Patent No. 7,127,379). Modeling analysis allows reliable predictions of the effects on cell growth of shifting the metabolism towards more efficient production of butadiene. One modeling method is the bilevel optimization approach, OptKnock (Burgard et al., *Biotechnol. Bioengineer.* 84:647-657 (2003)), which is applied to select gene knockouts that collectively result in better production of butadiene. Adaptive evolution also can be used to generate better producers of, for example, the but-3-en-1-ol intermediate or the butadiene product. Adaptive evolution is performed to improve both growth and production characteristics (Fong and Palsson, *Nat. Genet.* 36:1056-1058 (2004); Alper et al., *Science* 314:1565-1568 (2006)). Based on the results, subsequent rounds of modeling, genetic engineering and adaptive evolution can be applied to the butadiene producer to further increase production.

[00170] For large-scale production of butadiene, the above butadiene pathway-containing organism is cultured in a fermenter using a medium known in the art to support growth of the organism under anaerobic conditions. Fermentations are performed in either a batch, fed-batch or continuous manner. Anaerobic conditions are maintained by first sparging the medium with nitrogen and then sealing culture vessel (e.g., flasks can be sealed with a septum and crimp-cap). Microaerobic conditions also can be utilized by providing a small hole for limited aeration. The pH of the medium is maintained at a pH of 7 by addition of an acid, such as H2SO4. The growth rate is determined by measuring optical density using a spectrophotometer (600 nm), and the glucose uptake rate by monitoring carbon source depletion over time. Byproducts such as undesirable alcohols, organic acids, and residual glucose can be quantified by HPLC (Shimadzu) with an HPX-087 column (BioRad), using a refractive index detector for glucose and alcohols, and a UV detector for organic acids, Lin et al., *Biotechnol. Bioeng.*, 775-779 (2005).

EXAMPLE V

Pathway for the formation of butadiene precursor 3-buten-2-ol (but-2-en-1-ol) from acrylyl-CoA

[00171] This example describes pathways for converting acryly-CoA to 3-buten-2-ol, and further to butadiene. The conversion of acrylyl-CoA to 3-buten-2-ol is accomplished in four enzymatic steps. Acrylyl-CoA and acetyl-CoA are first condensed to 3-oxopent-4-enoyl-CoA by a beta-ketothiolase. The 3-oxopent-4-enoyl-CoA product is subsequently hydrolyzed to 3-oxopent-4-enoate by a CoA hydrolase, transferase or synthetase. Decarboxylation of the 3-ketoacid intermediate yields 3-buten-2-al, which is further reduced to 3-buten-2-ol by an alcohol dehydrogenase or ketone reductase.

[00172] Enzymes and gene candidates for catalyzing but-3-en-2-ol pathway reactions are described in further detail below.

[00173] Acrylyl-CoA and acetyl-CoA are condensed to form 3-oxopent-4-enoyl-CoA by a beta-ketothiolase (EC 2.3.1.16). Beta-ketothiolase enzymes catalyzing the formation of beta-ketovalerate from acetyl-CoA and propionyl-CoA are good candidates for catalyzing the formation of 3-oxopen-4-enoyl-CoA. *Zoogloea ramigera* possesses two ketothiolases that can form beta-ketovaleryl-CoA from propionyl-CoA and acetyl-CoA and *R. eutropha* has a beta-oxidation ketothiolase that is also capable of catalyzing this transformation (Gruys et al., U.S. Patent No. 5,958,745). The sequences of these genes or their translated proteins have not been reported, but several genes in *R. eutropha*, *Z. ramigera*, or other organisms can be identified based on sequence homology to *bktB* from *R. eutropha*.

Protein	GenBank ID	GI Number	Organism
phaA	YP_725941.1	113867452	*Ralstonia eutropha*
h16_A1713	YP_726205.1	113867716	*Ralstonia eutropha*
pcaF	YP_728366.1	116694155	*Ralstonia eutropha*
h16_B1369	YP_840888.1	116695312	*Ralstonia eutropha*
h16_A0170	YP_724690.1	113866201	*Ralstonia eutropha*
h16_A0462	YP_724980.1	113866491	*Ralstonia eutropha*
h16_A1528	YP_726028.1	113867539	*Ralstonia eutropha*
h16_B0381	YP_728545.1	116694334	*Ralstonia eutropha*
h16_B0662	YP_728824.1	116694613	*Ralstonia eutropha*
h16_B0759	YP_728921.1	116694710	*Ralstonia eutropha*
h16_B0668	YP_728830.1	116694619	*Ralstonia eutropha*
h16_A1720	YP_726212.1	113867723	*Ralstonia eutropha*
h16_A1887	YP_726356.1	113867867	*Ralstonia eutropha*
phbA	P07097.4	135759	*Zoogloea ramigera*
bktB	YP_002005382.1	194289475	*Cupriavidus taiwanensis*

Rmet_1362	YP_583514.1	94310304	*Ralstonia metallidurans*
Bphy_0975	YP_001857210.1	186475740	*Burkholderia phymatum*

[00174] Additional enzymes include beta-ketothiolases that are known to convert two molecules of acetyl-CoA into acetoacetyl-CoA (EC 2.1.3.9). Exemplary acetoacetyl-CoA thiolase enzymes include the gene products of *atoB* from *E. coli* (Martin et al., *Nat. Biotechnol.* 21:796-802 (2003)), *thlA and thlB* from *C. acetobutylicum* (Hanai et al., *Appl. Environ. Microbiol.* 73:7814-7818 (2007); Winzer et al., *J. Mol. Microbiol. Biotechnol.* 2:531-541 (2000)), and *ERG10* from *S. cerevisiae* (Hiser et al., *J. Biol. Chem.* 269:31383-31389 (1994)).

Protein	GenBank ID	GI Number	Organism
atoB	NP_416728	16130161	*Escherichia coli*
thlA	NP_349476.1	15896127	*Clostridium acetobutylicum*
thlB	NP_149242.1	15004782	*Clostridium acetobutylicum*
ERG10	NP_015297	6325229	*Saccharomyces cerevisiae*

[00175] Beta-ketoadipyl-CoA thiolase (EC 2.3.1.174), also called 3-oxoadipyl-CoA thiolase, converts beta-ketoadipyl-CoA to succinyl-CoA and acetyl-CoA, and is a key enzyme of the beta-ketoadipate pathway for aromatic compound degradation. The enzyme is widespread in soil bacteria and fungi including *Pseudomonas putida* (Harwood et al., *J. Bacteriol.* 176-6479-6488 (1994)) and *Acinetobacter calcoaceticus* (Doten et al., *J. Bacteriol.* 169:3168-3174 (1987)). The *P. putida* enzyme is a homotetramer bearing 45% sequence homology to beta-ketothiolases involved in PHB synthesis in *Ralstonia eutropha*, fatty acid degradation by human mitochondria and butyrate production by *Clostridium acetobutylicum* (Harwood et al., *supra*). A beta-ketoadipyl-CoA thiolase in *Pseudomonas knackmussii* (formerly *sp. B13*) has also been characterized (Gobel et al., *J. Bacteriol.* 184:216-223 (2002); Kaschabek et al., *supra*).

Protein	GenBank ID	GI Number	Organism
pcaF	NP_743536.1	506695	*Pseudomonas putida*
pcaF	AAC37148.1	141777	*Acinetobacter calcoaceticus*
catF	Q8VPF1.1	75404581	*Pseudomonas knackmussii*

[00176] Removal of the CoA moiety of 3-oxopent-4-enoyl-CoA product is catalyzed, for example, by 3-oxopent-4-enoyl-CoA hydrolase. The CoA hydrolase encoded by *acot12* from *Rattus norvegicus* brain (Robinson et al., *Biochem.Biophys.Res.Commun.*

71:959-965 (1976)) can react with several alternate substrates including butyryl-CoA, hexanoyl-CoA and malonyl-CoA. The human dicarboxylic acid thioesterase, encoded by *acot8,* exhibits activity on glutaryl-CoA, adipyl-CoA, suberyl-CoA, sebacyl-CoA, and dodecanedioyl-CoA (Westin et al., *J.Biol.Chem.* 280:38125-38132 (2005)). The closest *E. coli* homolog to this enzyme, *tesB*, can also hydrolyze a range of CoA thiolesters (Naggert et al., *J Biol Chem* 266:11044-11050 (1991)). A similar enzyme has also been characterized in the rat liver (Deana R., *Biochem Int* 26:767-773 (1992)). Additional enzymes with hydrolase activity in *E. coli* include *ybgC*, *paaI*, and *ybdB* (Kuznetsova, et al., *FEMS Microbiol Rev*, 2005, 29(2):263-279; Song et al., *J Biol Chem*, 2006, 281(16):11028-38). Though its sequence has not been reported, the enzyme from the mitochondrion of the pea leaf has a broad substrate specificity, with demonstrated activity on acetyl-CoA, propionyl-CoA, butyryl-CoA, palmitoyl-CoA, oleoyl-CoA, succinyl-CoA, and crotonyl-CoA (Zeiher et al., *Plant.Physiol.* 94:20-27 (1990)) The acetyl-CoA hydrolase, *ACH1*, from *S. cerevisiae* represents another candidate hydrolase (Buu et al., *J.Biol.Chem.* 278:17203-17209 (2003)) .

Protein	GenBank Accession #	GI#	Organism
acot12	NP_570103.1	18543355	*Rattus norvegicus*
tesB	NP_414986	16128437	*Escherichia coli*
acot8	CAA15502	3191970	*Homo sapiens*
acot8	NP_570112	51036669	*Rattus norvegicus*
tesA	NP_415027	16128478	*Escherichia coli*
ybgC	NP_415264	16128711	*Escherichia coli*
paaI	NP_415914	16129357	*Escherichia coli*
ybdB	NP_415129	16128580	*Escherichia coli*
ACH1	NP_009538	6319456	*Saccharomyces cerevisiae*

[00177] Additional hydrolase enzymes include 3-hydroxyisobutyryl-CoA hydrolase which has been described to efficiently catalyze the conversion of 3-hydroxyisobutyryl-CoA to 3-hydroxyisobutyrate during valine degradation (Shimomura et al., *J Biol Chem.* 269:14248-14253 (1994)). Genes encoding this enzyme include *hibch* of *Rattus norvegicus* (Shimomura et al., *Methods Enzymol.* 324:229-240 (2000)) and *Homo sapiens* (Shimomura et al., *supra*). Similar gene candidates can also be identified by sequence homology, including *hibch* of *Saccharomyces cerevisiae* and *BC_2292* of *Bacillus cereus*.

Protein	GenBank Accession #	GI#	Organism
hibch	Q5XIE6.2	146324906	*Rattus norvegicus*
hibch	Q6NVY1.2	146324905	*Homo sapiens*

Protein	GenBank Accession #	GI#	Organism
hibch	P28817.2	2506374	*Saccharomyces cerevisiae*
BC_2292	AP09256	29895975	*Bacillus cereus*

[00178] Decarboxylation of 3-oxopent-4-enoate is catalyzed by a 3-ketoacid decarboxylase. The acetoacetate decarboxylase (EC 4.1.1.4) from *Clostridium acetobutylicum,* encoded by *adc*, has a broad substrate specificity and has been shown to decarboxylate numerous alternate substrates including 2-ketocyclohexane carboxylate, 3-oxopentanoate, 2-oxo-3-phenylpropionic acid, 2-methyl-3-oxobutyrate and benzoyl-acetate (Rozzel et al., *J.Am.Chem.Soc.* 106:4937-4941 (1984); Benner and Rozzell, J.*Am.Chem.Soc. 103:993-994 (1981);* Autor et al.*, J Biol.Chem. 245:5214-5222 (1970)).* An acetoacetate decarboxylase has also been characterized in *Clostridium beijerinckii* (Ravagnani et al., *Mol.Microbiol* 37:1172-1185 (2000)). The acetoacetate decarboxylase from *Bacillus polymyxa*, characterized in cell-free extracts, also has a broad substrate specificity for 3-keto acids and can decarboxylate 3-oxopentanoate (Matiasek et al., *Curr.Microbiol* 42:276-281 (2001)). The gene encoding this enzyme has not been identified to date and the genome sequence of *B. polymyxa* is not yet available. Another *adc* is found in *Clostridium saccharoperbutylacetonicum* (Kosaka, et al., *Biosci.Biotechnol Biochem.* 71:58-68 (2007)). Additional gene candidates in other organisms, including *Clostridium botulinum* and *Bacillus amyloliquefaciens FZB42*, can be identified by sequence homology.

Protein	GenBank ID	GI No.	Organism
adc	NP_149328.1	15004868	*Clostridium acetobutylicum*
adc	AAP42566.1	31075386	*Clostridium saccharoperbutylacetonicum*
adc	YP_001310906.1	150018652	*Clostridium beijerinckii*
CLL_A2135	YP_001886324.1	187933144	*Clostridium botulinum*
RBAM_030030	YP_001422565.1	154687404	*Bacillus amyloliquefaciens*

[00179] Reduction of 3-buten-2-al to 3-buten-2-ol is catalyzed by an alcohol dehydrogenase or ketone reductase. Alcohol dehydrogenases described above in Example III are also suitable candidates for this transformation. There exist several exemplary alcohol dehydrogenases that convert a ketone to a hydroxyl functional group. Two such enzymes from *E. coli* are encoded by malate dehydrogenase (*mdh*) and lactate dehydrogenase (*ldhA*). In addition, lactate dehydrogenase from *Ralstonia eutropha* has been shown to demonstrate high activities on 2-ketoacids of various chain lengths

includings lactate, 2-oxobutyrate, 2-oxopentanoate and 2-oxoglutarate (Steinbuchel et al., *Eur.J.Biochem.* 130:329-334 (1983)). Conversion of alpha-ketoadipate into alpha-hydroxyadipate can be catalyzed by 2-ketoadipate reductase, an enzyme reported to be found in rat and in human placenta (Suda et al., *Arch.Biochem.Biophys.* 176:610-620 (1976); Suda et al., *Biochem.Biophys.Res.Commun.* 77:586-591 (1977)). An additional oxidoreductase is the mitochondrial 3-hydroxybutyrate dehydrogenase (*bdh*) from the human heart which has been cloned and characterized (Marks et al., *J.Biol.Chem.* 267:15459-15463 (1992)). Alcohol dehydrogenase enzymes of *C. beijerinckii* (Ismaiel et al., *J.Bacteriol.* 175:5097-5105 (1993)) and *T. brockii* (Lamed et al., *Biochem.J.* 195:183-190 (1981); Peretz et al., *Biochemistry.* 28:6549-6555 (1989)) convert acetone to isopropanol. Methyl ethyl ketone reductase catalyzes the reduction of MEK to 2-butanol. Exemplary MEK reductase enzymes can be found in *Rhodococcus ruber* (Kosjek et al., *Biotechnol Bioeng.* 86:55-62 (2004)) and *Pyrococcus furiosus* (van der Oost et al., *Eur.J.Biochem.* 268:3062-3068 (2001)).

Gene	GenBank Accession No.	GI No.	Organism
mdh	AAC76268.1	1789632	*Escherichia coli*
ldhA	NP_415898.1	16129341	*Escherichia coli*
ldh	YP_725182.1	113866693	*Ralstonia eutropha*
bdh	AAA58352.1	177198	*Homo sapiens*
adh	AAA23199.2	60592974	*Clostridium beijerinckii* NRRL B593
adh	P14941.1	113443	*Thermoanaerobacter brockii* HTD4
sadh	CAD36475	21615553	*Rhodococcus ruber*
adhA	AAC25556	3288810	*Pyrococcus furiosus*

[00180] Enzymes that catalyze the reduction of 3-oxobutanol to 1,3-butanediol are also applicable here. Such enzymes are found in organisms of the genus *Bacillus, Brevibacterium, Candida,* and *Klebsiella* among others, as described by Matsuyama et al. *J Mol Cat B Enz*, 11:513-521 (2001). One of these enzymes, SADH from *Candida parapsilosis*, was cloned and characterized in *E. coli*. A mutated *Rhodococcus* phenylacetaldehyde reductase (Sar268) and a *Leifonia* alcohol dehydrogenase have also been shown to catalyze this transformation at high yields (Itoh et al., *Appl.Microbiol Biotechnol.* 75:1249-1256 (2007)).

Gene	GenBank Accession No.	GI No.	Organism
sadh	BAA24528.1	2815409	*Candida parapsilosis*

EXAMPLE VI

Preparation of a Butadiene Producing Microbial Organism with a But-3-en-2-ol Pathway

[00181] This example describes the generation of a microbial organism capable of producing butadiene from pyruvate via a but-3-en-2-ol intermediate, in an organism engineered to have a butadiene pathway.

[00182] *Escherichia coli* is used as a target organism to engineer a butadiene-producing pathway. *E. coli* provides a good host for generating a non-naturally occurring microorganism capable of producing butadiene. *E. coli* is amenable to genetic manipulation and is known to be capable of producing various products, including ethanol, acetic acid, formic acid, lactic acid, and succinic acid, effectively under anaerobic or microaerobic conditions.

[00183] To generate an *E. coli* strain engineered to produce the butadiene pathway precursor, but-3-en-2-ol, a functional nucleic acids encoding the enzymes utilized in the pathway described in Example III, are expressed in *E. coli* using well known molecular biology techniques (see, for example, Sambrook, *supra*, 2001; Ausubel *supra*, 1999; Roberts et al., *supra*, 1989).

[00184] In particular, an *E. coli* strain is engineered to produce but-3-en-2-ol from acrylyl-CoA via the route described in Example III. For the first stage of pathway construction, genes encoding enzymes to transform acrylyl-CoA to but-3-en-2-ol are assembled onto a vector. The genes *phaA* (YP_725941.1), tesB (NP_414986), adc (NP_149328.1) and sadh (BAA24528.1) encoding beta-ketothiolase, 3-oxopent-4-enoyl-CoA hydrolase, 3-oxopent-4-enoate decarboxylase and 3-buten-2-one reductase, respectively, are cloned into the pZE13 vector (Expressys, Ruelzheim, Germany), under the control of the PA1/lacO promoter. The genes *mvk* (NP_357932.1), *mvaK2* (AAG02457.1) and, *ispS* (CAC35696.1) encoding alkyl phosphate kinase, alkyl diphosphate kinase and butadiene synthetase, respectively, are cloned into the pZA33 vector (Expressys, Ruelzheim, Germany) under the PA1/lacO promoter. The two plasmids are transformed into *E. coli* host strain containing $lacI^Q$, which allows inducible expression by addition of isopropyl-beta-D-1-thiogalactopyranoside (IPTG).

[00185] The resulting genetically engineered organism is cultured in glucose containing medium following procedures well known in the art (see, for example, Sambrook et al., *supra*, 2001). The expression of butadiene pathway genes is corroborated using methods well known in the art for determining polypeptide expression or enzymatic activity, including for example, Northern blots, PCR amplification of mRNA and immunoblotting. Enzymatic activities of the expressed enzymes are confirmed using assays specific for the individually activities. The ability of the engineered *E. coli* strain to produce butadiene is confirmed using HPLC, gas chromatography-mass spectrometry (GCMS) or liquid chromatography-mass spectrometry (LCMS).

[00186] Microbial strains engineered to have a functional butadiene synthesis pathway are further augmented by optimization for efficient utilization of the pathway. Briefly, the engineered strain is assessed to determine whether any of the exogenous genes are expressed at a rate limiting level. Expression is increased for any enzymes expressed at low levels that can limit the flux through the pathway by, for example, introduction of additional gene copy numbers. Strategies are also applied to improve production of butadiene precursor but-3-en-2-ol, such as mutagenesis, cloning and/or deletion of native genes involved in byproduct formation.

[00187] To generate better butadiene producers, metabolic modeling is utilized to optimize growth conditions. Modeling is also used to design gene knockouts that additionally optimize utilization of the pathway (see, for example, U.S. patent publications US 2002/0012939, US 2003/0224363, US 2004/0029149, US 2004/0072723, US 2003/0059792, US 2002/0168654 and US 2004/0009466, and in U.S. Patent No. 7,127,379). Modeling analysis allows reliable predictions of the effects on cell growth of shifting the metabolism towards more efficient production of butadiene. One modeling method is the bilevel optimization approach, OptKnock (Burgard et al., *Biotechnol. Bioengineer.* 84:647-657 (2003)), which is applied to select gene knockouts that collectively result in better production of butadiene. Adaptive evolution also can be used to generate better producers of, for example, the but-3-en-2-ol intermediate or the butadiene product. Adaptive evolution is performed to improve both growth and production characteristics (Fong and Palsson, *Nat. Genet.* 36:1056-1058 (2004); Alper et al., *Science* 314:1565-1568 (2006)). Based on the results, subsequent rounds of modeling, genetic engineering and adaptive evolution can be applied to the butadiene producer to further increase production.

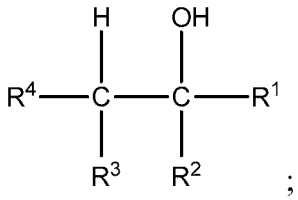
[00188] For large-scale production of butadiene, the above butadiene pathway-containing organism is cultured in a fermenter using a medium known in the art to support growth of the organism under anaerobic conditions. Fermentations are performed in either a batch, fed-batch or continuous manner. Anaerobic conditions are maintained by first sparging the medium with nitrogen and then sealing culture vessel (e.g., flasks can be sealed with a septum and crimp-cap). Microaerobic conditions also can be utilized by providing a small hole for limited aeration. The pH of the medium is maintained at a pH of 7 by addition of an acid, such as H2SO4. The growth rate is determined by measuring optical density using a spectrophotometer (600 nm), and the glucose uptake rate by monitoring carbon source depletion over time. Byproducts such as undesirable alcohols, organic acids, and residual glucose can be quantified by HPLC (Shimadzu) with an HPX-087 column (BioRad), using a refractive index detector for glucose and alcohols, and a UV detector for organic acids, Lin et al., *Biotechnol. Bioeng.*, 775-779 (2005).

[00189] Throughout this application various publications have been referenced. The disclosures of these publications in their entireties, including GenBank and GI number publications, are hereby incorporated by reference in this application in order to more fully describe the state of the art to which this invention pertains. Although the invention has been described with reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

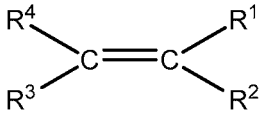
1. A non-naturally occurring microbial organism, said microbial organism having an alkene pathway and comprising at least one exogenous nucleic acid encoding an alkene pathway enzyme expressed in a sufficient amount to convert an alcohol to an alkene, wherein said alkene pathway comprises a pathway selected from: (1) an alcohol kinase and a phosphate lyase; (2) a diphosphokinase and a diphosphate lyase; and (3) an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase,

wherein said alcohol is a compound of Formula (I)



;

and wherein said alkene is a compound of Formula (II)



wherein R^1, R^2, R^3, and R^4 are independently the same between said alcohol and said alkene and are each independently (a) hydrogen, cyano, halo, or nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more substituents Q; or (c) –$C(O)R^{1a}$, –$C(O)OR^{1a}$, –$C(O)NR^{1b}R^{1c}$, –$C(NR^{1a})NR^{1b}R^{1c}$, –OR^{1a}, –$OC(O)R^{1a}$, –$OC(O)OR^{1a}$, –$OC(O)NR^{1b}R^{1c}$, –$OC(=NR^{1a})NR^{1b}R^{1c}$, –$OS(O)R^{1a}$, –$OS(O)_2R^{1a}$, –$OS(O)NR^{1b}R^{1c}$, –$OS(O)_2NR^{1b}R^{1c}$, –$NR^{1b}R^{1c}$, –$NR^{1a}C(O)R^{1d}$, –$NR^{1a}C(O)OR^{1d}$, –$NR^{1a}C(O)NR^{1b}R^{1c}$, –$NR^{1a}C(=NR^{1d})NR^{1b}R^{1c}$, –$NR^{1a}S(O)R^{1d}$, –$NR^{1a}S(O)_2R^{1d}$, –$NR^{1a}S(O)NR^{1b}R^{1c}$, –$NR^{1a}S(O)_2NR^{1b}R^{1c}$, –$SR^{1a}$, –$S(O)R^{1a}$, –$S(O)_2R^{1a}$, –$S(O)NR^{1b}R^{1c}$, or –$S(O)_2NR^{1b}R^{1c}$; wherein each R^{1a}, R^{1b}, R^{1c}, and R^{1d} is independently hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or R^{1a} and R^{1c} together with the C and N atoms to which they are attached form heterocyclyl; or R^{1b} and R^{1c} together with the N atom to which they are attached form heterocyclyl;

wherein each Q is independently selected from (a) oxo, cyano, halo, and nitro;

(b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl, each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (c) $-C(O)R^a$, $-C(O)OR^a$, $-C(O)NR^bR^c$, $-C(NR^a)NR^bR^c$, $-OR^a$, $-OC(O)R^a$, $-OC(O)OR^a$, $-OC(O)NR^bR^c$, $-OC(=NR^a)NR^bR^c$, $-OS(O)R^a$, $-OS(O)_2R^a$, $-OS(O)NR^bR^c$, $-OS(O)_2NR^bR^c$, $-NR^bR^c$, $-NR^aC(O)R^d$, $-NR^aC(O)OR^d$, $-NR^aC(O)NR^bR^c$, $-NR^aC(=NR^d)NR^bR^c$, $-NR^aS(O)R^d$, $-NR^aS(O)_2R^d$, $-NR^aS(O)NR^bR^c$, $-NR^aS(O)_2NR^bR^c$, $-SR^a$, $-S(O)R^a$, $-S(O)_2R^a$, $-S(O)NR^bR^c$, and $-S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heterocyclyl, optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a;

wherein each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl; and (c) $-C(O)R^e$, $-C(O)OR^e$, $-C(O)NR^fR^g$, $-C(NR^e)NR^fR^g$, $-OR^e$, $-OC(O)R^e$, $-OC(O)OR^e$, $-OC(O)NR^fR^g$, $-OC(=NR^e)NR^fR^g$, $-OS(O)R^e$, $-OS(O)_2R^e$, $-OS(O)NR^fR^g$, $-OS(O)_2NR^fR^g$, $-NR^fR^g$, $-NR^eC(O)R^h$, $-NR^eC(O)OR^f$, $-NR^eC(O)NR^fR^g$, $-NR^eC(=NR^h)NR^fR^g$, $-NR^eS(O)R^h$, $-NR^eS(O)_2R^h$, $-NR^eS(O)NR^fR^g$, $-NR^eS(O)_2NR^fR^g$, $-SR^e$, $-S(O)R^e$, $-S(O)_2R^e$, $-S(O)NR^fR^g$, and $-S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heterocyclyl.

2. The non-naturally occurring microbial organism of claim 1, wherein said microbial organism comprises two or three exogenous nucleic acids each encoding an alkene pathway enzyme.

3. The non-naturally occurring microbial organism of claim 2, wherein said two exogenous nucleic acids encode an alcohol kinase and a phosphate lyase.

4. The non-naturally occurring microbial organism of claim 2, wherein said two exogenous nucleic acids encode a diphosphokinase and a diphosphate lyase.

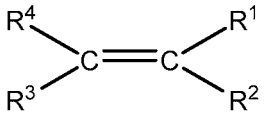
5. The non-naturally occurring microbial organism of claim 2, wherein said three exogenous nucleic acids encode an alcohol kinase, an alkyl phosphate kinase and a diphosphate lyase.

6. The non-naturally occurring microbial organism of claim 1, wherein said at least one exogenous nucleic acid is a heterologous nucleic acid.

7. The non-naturally occurring microbial organism of claim 1, wherein said non-naturally occurring microbial organism is in a substantially anaerobic culture medium.

8. The non-naturally occurring microbial organism of claim 1, wherein said microbial organism converts ethanol to ethylene, n-propanol to propylene, isopropanol to propylene, n-butanol to but-1-ene, isobutanol to isobutylene, tert-butanol to isobutylene, butan-2-ol to but-1-ene or but-2-ene, pentan-1-ol to pent-1-ene, 3-methylbutan-1-ol to 3-methylbut-1-ene, pentan-2-ol to pent-2-ene, pental-3-ol to pent-2-ene, 2-methylbutan-1-ol to 2-methylbut-1-ene, 3-methylbutan-2-ol to 3-methylbut-1-ene, 2-methylbutan-2-ol to 2-methylbut-1-ene or 2-methylbut-2-ene, 3-methylbut-3-en-1-ol to isoprene, 2-methylbut-3-en-2-ol to isoprene, 2-methylbut-3-en-2-ol to 3-methylbuta-1,2-diene, 2-methylbut-3-en-1-ol to isoprene, 3-methylbut-3-en-2-ol to isoprene, but-3-en-1-ol to 1,3-butadiene, but-3-en-2-ol to 1,3-butadiene, 1-phenylethanol to styrene, 2-phenylethanol to styrene, dimethylallyl alcohol to isoprene, or but-2-en-1-ol to 1,3-butadiene.

9. A method for producing an alkene comprising culturing the non-naturally occurring microbial organism of any one of claims 1-7 under conditions and for a sufficient period of time to produce said alkene, wherein said alkene is a compound of Formula (II)



wherein R^1, R^2, R^3, and R^4 are each independently (a) hydrogen, cyano, halo, or nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more substituents Q; or (c) $-C(O)R^{1a}$, $-C(O)OR^{1a}$, $-C(O)NR^{1b}R^{1c}$, $-C(NR^{1a})NR^{1b}R^{1c}$, $-OR^{1a}$, $-OC(O)R^{1a}$, $-OC(O)OR^{1a}$, $-OC(O)NR^{1b}R^{1c}$, $-OC(=NR^{1a})NR^{1b}R^{1c}$, $-OS(O)R^{1a}$, $-OS(O)_2R^{1a}$, $-OS(O)NR^{1b}R^{1c}$, $-OS(O)_2NR^{1b}R^{1c}$, $-NR^{1b}R^{1c}$, $-NR^{1a}C(O)R^{1d}$, $-NR^{1a}C(O)OR^{1d}$,

$-NR^{1a}C(O)NR^{1b}R^{1c}$, $-NR^{1a}C(=NR^{1d})NR^{1b}R^{1c}$, $-NR^{1a}S(O)R^{1d}$, $-NR^{1a}S(O)_2R^{1d}$, $-NR^{1a}S(O)NR^{1b}R^{1c}$, $-NR^{1a}S(O)_2NR^{1b}R^{1c}$, $-SR^{1a}$, $-S(O)R^{1a}$, $-S(O)_2R^{1a}$, $-S(O)NR^{1b}R^{1c}$, or $-S(O)_2NR^{1b}R^{1c}$; wherein each R^{1a}, R^{1b}, R^{1c}, and R^{1d} is independently hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or R^{1a} and R^{1c} together with the C and N atoms to which they are attached form heterocyclyl; or R^{1b} and R^{1c} together with the N atom to which they are attached form heterocyclyl;

wherein each Q is independently selected from (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl, each of which is further optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; and (c) $-C(O)R^a$, $-C(O)OR^a$, $-C(O)NR^bR^c$, $-C(NR^a)NR^bR^c$, $-OR^a$, $-OC(O)R^a$, $-OC(O)OR^a$, $-OC(O)NR^bR^c$, $-OC(=NR^a)NR^bR^c$, $-OS(O)R^a$, $-OS(O)_2R^a$, $-OS(O)NR^bR^c$, $-OS(O)_2NR^bR^c$, $-NR^bR^c$, $-NR^aC(O)R^d$, $-NR^aC(O)OR^d$, $-NR^aC(O)NR^bR^c$, $-NR^aC(=NR^d)NR^bR^c$, $-NR^aS(O)R^d$, $-NR^aS(O)_2R^d$, $-NR^aS(O)NR^bR^c$, $-NR^aS(O)_2NR^bR^c$, $-SR^a$, $-S(O)R^a$, $-S(O)_2R^a$, $-S(O)NR^bR^c$, and $-S(O)_2NR^bR^c$, wherein each R^a, R^b, R^c, and R^d is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl, each optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a; or (iii) R^b and R^c together with the N atom to which they are attached form heterocyclyl, optionally substituted with one or more, in one embodiment, one, two, three, or four, substituents Q^a;

wherein each Q^a is independently selected from the group consisting of (a) oxo, cyano, halo, and nitro; (b) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, and heterocyclyl; and (c) $-C(O)R^e$, $-C(O)OR^e$, $-C(O)NR^fR^g$, $-C(NR^e)NR^fR^g$, $-OR^e$, $-OC(O)R^e$, $-OC(O)OR^e$, $-OC(O)NR^fR^g$, $-OC(=NR^e)NR^fR^g$, $-OS(O)R^e$, $-OS(O)_2R^e$, $-OS(O)NR^fR^g$, $-OS(O)_2NR^fR^g$, $-NR^fR^g$, $-NR^eC(O)R^h$, $-NR^eC(O)OR^f$, $-NR^eC(O)NR^fR^g$, $-NR^eC(=NR^h)NR^fR^g$, $-NR^eS(O)R^h$, $-NR^eS(O)_2R^h$, $-NR^eS(O)NR^fR^g$, $-NR^eS(O)_2NR^fR^g$, $-SR^e$, $-S(O)R^e$, $-S(O)_2R^e$, $-S(O)NR^fR^g$, and $-S(O)_2NR^fR^g$; wherein each R^e, R^f, R^g, and R^h is independently (i) hydrogen; (ii) C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} cycloalkyl, C_{6-14} aryl, C_{7-15} aralkyl, heteroaryl, or heterocyclyl; or (iii) R^f and R^g together with the N atom to which they are attached form heterocyclyl.

10. The method of claim 9, wherein said alkene is selected from Ethylene, Propylene, Propylene, But-1-ene, Isobutylene, Isobutylene, But-1-ene, but-2-ene, Pent-1-ene, 3-methylbut-1-ene, Pent-2-ene, 2-methylbut-1-ene, 3-Methylbut-1-ene, 2-Methylbut-1-ene, 2-Methylbut-2-ene, Isoprene, 3-Methylbuta-1,2-diene, 1,3-Butadiene and Styrene.

5