METHODS AND COMPOSITIONS FOR PRODUCING OLEFINS

Applicant: REG LIFE SCIENCES, LLC, Ames, IA (US)

Inventors: Murtaza F. Alibhai, Austin, TX (US); Mathew A. Rude, Redwood City, CA (US); Andreas W. Schirmer, South San Francisco, CA (US)

Related U.S. Application Data

Continuation of application No. 13/452,449, filed on Apr. 20, 2012, now Pat. No. 8,597,922, which is a division of application No. 12/673,752, filed on Apr. 14, 2011, now Pat. No. 8,183,028, filed as application No. PCT/US2008/014029 on Dec. 22, 2008.

Compositions and methods for producing olefins are described herein. The olefins can be used to produce biofuels.
FIG. 1A

FIG. 1B

TIC: as3.D\data.ms

TIC: as-1.D\data.ms

hexacosane
Internal standard

hexacosane
Internal standard

i-C20

a-C20 olefin

i-C19 olefin

n-C19 olefin

n-C21

n-C21

i-C19

i-C21

n-C19

n-C20

i-C20

n-C19

i-C19

n-C21

n-C20

i-C20 olefin
FIG. 4A

1-nonadecene

TIC: as3.D\data.ms
TIC: as6.D\data.ms

eicosanoic acid

hexacosane (internal standard)

FIG. 4B

1-heptadecene

1-pentadecene

1-nonadecene

hexacosane (internal standard)

stearic acid
eicosanoic acid

TIC: as001.D\data.ms
TIC: as007.D\data.ms
**Jeotgalicoccus sp. ATCC8456 orf880 (SEQ ID NO:1)**

60  ATGGCAACA C TTAAGGGAGA TAAAGGCTATA GATAATACTG TGAAAGTATT AAAGCAAGGT
120  CTGATACCGT ATGTCAGCTGA AGGGCGCTAGA CCAAACGTTA TTGTTAATAAC GCTTTTGGT
180  AGTATTGGAT GATACGTTGA ATGTGGCITG GTAATCAAAA ATGTTCGCGA AGTTGTAGAC
240  AGTAACAAA CACGTCGCGA AGGTCCGGAG CAACCGTTGA AGAAGATTTT
300  AAGGTTACA AGAGCAAGAT GC CGGGTTAGT AGGAGCTGTT AGAAAGTTTT AGAAGAAGCA
360  ATTATGGAGA CTGGTAAAGG GGATATTTCTG TCCGCTGGAG GTACACCGACT TTACGATTIT
420  GGACATTGGG AAGACTACCTT AGGTAAACCA ATGGACTCAA GAACCTGTGC GATGGACTTA
480  ATGGGAATAC TCCGCCCATG AAGATTACAC AGCATGAACTG TTTCACTGG GTTTACACGC
540  ATGACGCAAA ACCAATGCAC AGCTGAAAAA ATGAAATCTG AACCTGACTA TCCGATATAA
600  TGCCTCAAG AGGTCTCTTG TTACTATCCA TCCGTCTCAG TCCCGACGG TAAAAGCAA
660  GTAGACATTG CCAGAGGTTT CGGATCAATTT CCAGACGGGG TAGACCTGCT TACGATTTT
720  TATGGTACAA CACATGATGA ATCCTTTTGG GAGATATCCA ATGAATTCGG CCAGAGGAGA
780  TCCGAAATCTT GCAAGAGATC ACCACTCTAT CTATTCTCCG AGGCTTGGTG AGATTCTCGG
840  ACAATATCAA GTTGCTCGAG TGATGCGGAT ACAGTAAATC TGATGAGAGA AAGATATGAA
900  TACATTGCGG CCAGACCAAC ATGTGGAGCT GACTTAAACAC
960  CGACAGCTAA

**FIG. 7A**

**Jeotgalicoccus sp. ATCC8456 orf880 (SEQ ID NO:2)**

60  MATLKRDKGL DNTLYNKRNR LNTSVPQTKA LGGKPFVVT GKE3A5MFYN
120  NDVQQREGM PKNIVWLFPG KGAITVHDVG KGVRKCALPM SLMTENLNY VRELTRLTH
180  ANTQAESEMD EBVUIYKSEV LUTKVIRWA GVQAPHEDE RIAINDMDIM DISPRALGQAF
240  KGYQASKER KTVREDWLEQ I1ETRGNHI PFRGTAHLEF AHWEDYLGNP MDSSRTCIAIDL
300  MNTRFRLLAI NRPVFSFLHA MNERNPTREK IKSEPDPYAYK FAQERVRYYP FWPPFLPGKAK
360  VDIDFGQVTH PAVQGLLALDV YGTTHDESLE DPENFRFR FRWQDSPFD LIPQQCDYTW
420  TNHRCAEWEI TVIJIMEEMK YPAEKITYDV PEQDLEVDLN SIPGVKGSF VIKNVRBBVD
430  RT

**FIG. 7B**
Jeotgalicoccus sp. ATCC8456 16s rRNA (partial sequence) (SEQ ID NO:3)

```
60  GGTGCTCTGGA TTAGAATCTTTCACCCCTATGA TCACAGATGACGTTTGGATTGCTTCCTAGTTTTGCACTG
120  AGAGATTCAACCGGCCTCTCGAAGTTTACCGTTTGACCAACGCGCTACGAGCAGCGTICA TCCTGAGCCA
180  GGTGTCAGCG TCAAGAACCACTGCTGGATTATTGCCAGAACGCTGCTTCAAGACGCTGCTGGATTGCTAC
240  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
300  GGTGTCAGCG TCAAGAACCACTGCTGGATTATTGCCAGAACGCTGCTTCAAGACGCTGCTGGATTGCTAC
360  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
420  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
480  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
540  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
600  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
660  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
720  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
780  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
840  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
900  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
960  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1020  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1080  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1140  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1200  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1260  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1320  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1380  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1440  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1500  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT
1550  GTGAGCGGCTACGTTGGGTAAGTTCGAGAGCTCCCTGCTATTACGCTGGTTAAGTTCGAGAGCTCCCTGCTATTACGCT

FIG. 7C
Jeotgalicoccus sp. ATCC8456 orf880, codon-optimized DNA (SEQ ID NO:4)

1  ATGCCTACTC TGAACGGTGCA CAAAGGCTTG GATAAACACTC TGAAAGTCTT GAAACAGAGT
61  TACCTGCTAC ATACCAACCA CGGCAACAGGG CTGAACGCCA GGGTCTCTGA AACACAAAGCC
121  CGCGGTGCGA AACGCTGCTG GTTTTGACCC GGGCAAAAGG GGGCAGGAGAT GTTCTTTAAGCT
181  AACGCTGCGG TGCGCGCGGG GGCGAGCTGG CAGAAAATGA TGTGTTGGTC CGTGTTTGGC
241  AACGGTGCGA TCCCTACCGTT GCGGCGCGAG CAAACCTTAG ACCCTTAAAAGG ACTGGTTGAT
301  TTCTGTGTTGA CTGAGGGCAGA CCTGAAATCTG GTACGCGAAC TGACCCTGCAC CGTGCTGGCT
361  GCGAAACCGC ACGCTATYCA AATCTACGCTG AGGCGTACGA CGCTACGGTA AAGCAGCTGT
421  CTGCCTGAAAG AGTGGGCGCA CGCTGCGCACA GGTTGTCGAG CAGGCGCAGA GCAAGTTGAG
481  CGATCTCTCA CGGATTATGA TTTATGATCT GATAGCTTCC GTGCTCTGGG TGGCCCGATT
541  ATGACCGAAA CCGGTAAGAG TAACTACACG CCGACCAGAG GGAACCGCTT GTCAGATTTT
601  GCGACCTGGG AGGATTATCTG CGGTAATCGA ACGGACCTTC GTACCTGCGC GATCGATGCTG
661  ATGAAACAGGT TTGCGCGCTCT GATCGCTATC ACGGCCTTTC TTCCCTTGGA TTCCCTGGCG
721  ATGAAACAGGA ACCGATCACG TCTGGAGAGG ATTAGTCCG AGGCGAATGA CCGTAACAAA
781  TTGCGACAGG AGTGGCGCTG CTACACCCCG TTGCTTTCTT TCTGGCGGCT TAAAGCGAAGG
841  GCGACTATGG ACTCGACGGG CTTCCACCCT GTAGACCCCT CCGGCGGTTG TGGCCCGATG
901  TACGGTACCA CCGGATGAGA AGGCGCGTGG GATGACGCGA AGGAATTTGC CCCGGAGCGT
961  TTGGAGACCT GGCGGGGCTC TCCATCTCAG CGTATGCGCG GGGGCGGGGC TGATCACGGG
1021  ACCCATACCC GGGCGCGCGG CGAGCGAGATC CGCGTCAGTCC GATATGACAGA ACGAACGAAA
1141  TACCTGCGCG AGGAAATACGC TTATGGACCT CGCAGACAGG ACGTGAGATG AAGACGACAC
1201  TCCATCCCGG GCTATGCTAA AACGCGCCTTT GTTATCAGAA ACGTCGCGGA AGATACGAT
1261  CGACCCAAA

FIG. 7D

Jeotgalicoccus sp. ATCC8456 orf880, Protein (SEQ ID NO:5)

1  MATLKLHKGKL DNLKLKVLQG YLTXTNOQRNR LNTSVPQTGA LGGKPFVVVT GKEGAEFFYN
61  NDDYQIQGML PRRKVNLFGL KAGIHTDYGK KHDVXALF SLMIEGNY NRELDAVIIH
121  ANDORWEAMD EGNYRBEIV LITKVSRWA GVQAPPEDIE RIATIMDIIMI DESRALGIGAF
181  KGYASKAEAR KVEVDELEFO LESDFGKNNHQ PEGZALYEF AHMEDLQAG NDSQXCAIIDL
241  MNTRPLTIAE NPRVSFGILHA MNENITREK IKSSEPDAAYK PAVEVRAKYP FPFLPGKAK
301  VDIDPGYTVT FAVGGILADV GTTHTDESLN DDNPERPER FETWDOGSPF LIRQOOGGDDW
361  TNRHCGNWI TVIAMEETMK FYAEKCYTV PEQLESVDLN STIHYVKSSG FIKVREWVD
421  RT

FIG. 7E
**Corynebacterium efficiens orf_CE2459 (NP_739069) DNA (SEQ ID NO:6)**

1 TCAGCCGTCCC AGCGCAGCCT GCATGCGCGG TCGGAAGCTT GTCTGACGGG
61 GAACCGGCTA TCAGCCGGAT GCCTGCTAGA CACCAAGACCG GCAGGGGATA AAGCCTTCAC
121 GTGCGGCTGGA AGGGGGCCCA TGCGGATCTC TCGGCGGCTG CAGCGGCTGAC CGGGTGACAC
181 GGGCGCCCTCC CAGAACGCTGT AGACGCTCCT TAGCTGACCG GGGGCGCCCAG
241 GTGCTCCGCGG GACAGAAGAG GTCGCGCTTG AAACGCGACT GCGTCTCCAC ACTGACTTGGG
301 GTGCGGCTTGG TGGCGGCTGGA TCGCGGCTGGA CGCGGGCCAC CAGGGCGGAGCG
361 CAGGGCTGCGG TGGCGGCTTGG CAGGCTGCTG GCGACGGGAG GATGAGCGGG
421 GGGCGGCGAGG TGGCGGCGAGG CGCGGCGGACG CCGCGGCGGAC CGCGGCGGACG
481 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
541 GGGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
601 GGGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
661 GGGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
721 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
781 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
841 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
901 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
961 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
1021 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
1081 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
1141 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
1201 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG
1261 CCGCGGCGGACG TGGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG CCGCGGCGGACG

**FIG. 7F**

**Corynebacterium efficiens orf_CE2459 (NP_739069) Protein (SEQ ID NO:7)**

1 MEVEFFMGMQI SSCPPARIEQ APNLLRHCYYL FLSLRRKAG ISPDANTPLR SRMLFKPVIT
61 VRGQAGYELF YNDRMRRGKD AMPAVRHYML FGBGAHVSDL GEHEHRLRNGQ LADVAYDDDK
121 VARLPGDMVR FVYRVRVQQAMA REETPTYDGA ALAEOGRAAYR WAGELQSQKAE ASRRAHIAEM
181 LVYQFCNHPLK GHALGWINRA RLNRWALKLI QRQARAGERHV APGSALEAMS RLVGPQDFGELV
241 DASIAHTEQ LIYRTRPAVAV VIATPFAGSSAVL VEHPWSVEKI RSGGQPVFAA FAGQVPVRYP
301 FYMLPAIJAT TDYIEICCPYV HEGEYTVDI IDHGINIDPNEV EPNSAFQPER FLSDREDLSTQ
361 EYERLTSFV PQGGAVVYTG HRCFOEKTAM AALITAAMEAL CRPVVVLSTD PAVTRFPWTQ
421 MLTRSETGMR VRVER

**FIG. 7G**
Kokuria rhizophila orf_KRH21570 (YP_001856010)DNA (SEQ ID NO:8)

1 ATGACTTCCAC CTTCGTTGCA GACCGTGTTG CGACCGAGCC CGTCTCTACT CGGCTCGGGG
61 TACCTCTTTTG CCTCCGCCCC ACGAACGCCC GCGGCGCTCT CTCGCGACTC GGGTGGGGCC
121 GTCCGCGTGC CTCGCGTGGG CTTACGGAGC CTCCCTCTCC CGCGAGGCGA GGGCTCCAGG
181 CTCTTTCTAGG ACACCTCGGG CGTGCGGCGC GACGCCGGCA TGGCGCGAGT CTGCGAGGCG
241 CGCTCTCTTG GTGCCGCGGGC CTTGACAAGG CTTGACCGGG AGCGCGCCAG TGGCGCCCGG
301 AACAACATTG CGCGAAGCGA CTACAGGAGG GACGCGCTGG GGGGCTCAA CCGGCGCGTG
361 GCGGACGGAG CGAGAGACTG CGTGCGAGGG TGACAGGACG ACAGTACACG CTAGAAGAAC
421 ACGGCGATCG CTTCCGCGGG CCGTTCCCTC CGTGAGGGCG CTCGGCGATG GGGCGCGGCG
481 GACGNTCACT CGCGGCGCGG CGCGAGTGGG CGCTGCGCTG ACGCCCTCGG CGCGCGCGCG
541 ACSCACCGTG TGCCCGGACT CGACCGGATC CGCGCGCTCC GCGCGTCCTC GCGGCCTATG
601 AAGCGACGTC GCGCGGCGAG GTCCGACCGA CGCGCGCTAC CGCGCGCGCG CCGCGCGAGC
661 GCGCGCGCTG CGACGCGACG CGCGCGCTGG CGACGAGCGA CGCGCGCGAT CGCGCGCGAG
721 AACCTGACCC GCGCTGGGCG CCGGTCCCACT CGTCCGCGGG CGCGCGCGCG CCGCGCGCGG
781 GTGCCGACG CGCGCTGCGC CGCGCGCGCG CGCGATCGGC CGCGCGCGCG CGCGCGCGCGG
841 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
901 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
961 ACGGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
1021 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
1081 ACGGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
1141 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
1201 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG
1261 CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCG CGCGCGCGCGG

FIG. 7H
Kokuria rhizophila orf KRH21570 (YP_001856010) codon-optimized DNA (SEQ ID NO:9)

1 ATGAGAGGCC GTTTCGCGCA GACCCGATGC GACAGGGCC CGACCTGGCT GCCTTGGGCTG
61 TACTTGGTGC GAAACCGGTC TCCTCGGCTG GCCTGCTGTC GCAAGCGTGAG GCTGTCGCTCA
121 GTGCTGCGATGC GTGCTGCGAG TAAGCAGACG TGTCGTCGAG GCCTGAGGGCA AGCGGCCGAA
181 CTTCTTGCTG CTGAGCTGCTG GCCTCGCGCTG AAGGCGGCGG GCCTTGGCGAG
241 GCCTCGCTCG GCCTCGCTCG GTGGACCGGTG TCGAAGGATG TCAAGAAGCT CTACGGACGC
301 AAGCGACCTGG GTCCGGTAAA AGTGAACGCT CCGAGAACA GGCTCGTGCG CCCCATGGCA
361 GCTGGTGGCTG CGAAGAGTGG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
421 GAGCGAGTGTC GCCGAGAGTG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
481 GCCTCGCGCG GCCGAGAGTG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
541 AGCCACCCCTG TGCCGGTAAA GAGCTGCCCT GCCCGAGGTT GCCGAGAGTG TCGAGAGCTG
601 AAGCGAGTGTC GCCGAGAGTG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
661 GCTGGTGGCTG CGAAGAGTGG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
721 AATTTTGCCGC GCCGAGAGTG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
781 GCTGGTGGCTG CGAAGAGTGG TCGAGAGCTG GCTGGTGGCTG TCAAGAAGCT CTACGGACGC
841 TTGGAGGGCTG TCAGCGAGGTG AGCAGGAGGG TCAGCGAGGTG AGCAGGAGGG TCAGCGAGGTG
901 GTCTGGGAGG TGCCGGCGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
961 AAGGGTTGACGC GCTGGGAGG TGCCGGCGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
1021 CGTGGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
1081 ACCTTTTACC GCCGGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
1141 AGCTGGGAGG TGCCGGCGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
1201 CGTGGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG
1261 GCCTGGGAGG TGCCGGCGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG GCCGGAGGCG

FIG. 7I

Kokuria rhizophila orf KRH21570 (YP_001856010) Protein (SEQ ID NO:10)

1 MTSPFQGIRS EQGASLLELSS VLFGASKARR AGLSDSGSCP VEMPLLGKQTV VLVRGEEBGVK
61 LFVLYTSGGVR DGAMCGCVQG DPLGAGAVHG LODERARHVK NQADAMYED ERVAVKYPFV
121 AELLENLVAR WIKOMYVYOS TAJACGFRSF RWAGLQGEVYP EMDFAARRMS RLLDTEFSGA
181 TSLWRDRDRR AIADRRFAALI KOVRAKVKNA FEDSVLHMA AMIDEHGLV DAKTAGELEQ
241 MRRPNRPNVAA RAFFAAAATL RHEHSREDI RAASECRGOST LLLDVEFAVAF AQEVRRFYPF
301 EVMPLAEVQ DTSIQQCQPVH KGERLVVDIL GINMTFFSWD RAAFEEDPERF LGVEADARAT
361 TQPQOOGAEV RTGHRCFCEK LAITSSLGAAV VALCRRKSLV PQDQDLLIFS WTHEMLTRPVF
421 GVRVRTRR

FIG. 7J
Methyllobacterium populi orf Mpop1292 (YP_001923998)codon-optimized DNA (SEQ ID NO:11)

1 ATGCGGCGTG CGATGCGCAC CGACCTTTTC CGCCAGCTCC GCACCCCTCG GCGTGAGGCA
61 GCTTGACAAT GCAGCGCGGC CGCCGGCGAC GGGGCGATAC GCTTCACTGC TACCGTTGCT
121 CGCGTGATCC ACGCGACGCC CTGCGCGCCG CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
181 TGGCGCCGAG CCGCGACGAC CGCCGGCCCG GCACCGAGGT CAGCGGCTAC CGCGTGATCC
241 CGCGTGATCC ACGCGACGCC CTGCGCGCCG CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
301 GCCCGCAGCA CCTGGAGTTA ACGTGCAATTC CTGAGCGCGC CGCCGGCCCG CAGCGGCTAC
361 GCTTGCCGAC AGCGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
421 CGCGTGATCC ACGCGACGCC CTGCGCGCCG CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
481 GCGCGGCGTG AGCGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
541 ATGGTGGCCG ACAGCGCGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
601 GCACCGAGGT CGCGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
661 CGCGTGATCC ACGCGACGCC CTGCGCGCCG CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
721 ATGGTGGCCG AGTGGTGGCCG GCACCGAGGT CGCGCGACGCC CGCCGGCCCG CAGCGGCTAC
781 ATGGTGGCCG AGTGGTGGCCG GCACCGAGGT CGCGCGACGCC CGCCGGCCCG CAGCGGCTAC
841 CCTGCGCGGCG CGCGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
901 TTATCGCGCG CAGCGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
961 GAATGGTGCC CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
1021 CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
1081 CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
1141 CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
1201 CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC
1261 CGCGGCGACGCC CGCCGGCCCG CAGCGGCTAC CGCGTGATCC

FIG. 7K

Methyllobacterium populi orf Mpop1292 (YP_001923998)Protein (SEQ ID NO:12)

1 MPAAATAHFR RXAARLRFEP APOSTALLLRE EYGGFIIRRC RPHSDDDFFAA RLLLSFVCD
61 SGABAAHFLY DCHRFTSRHA LFTISPFLAQ DMSGVWLDG AAILAKRMRF LSLVSGFAQL
121 RLAGAERAAN RAPWVSMARK DTVVVLIDEGH AVLTVLACG WGGVLGIPG DVARERFAAM
181 IDCTGAVGDR NNORHLAR TRERLKWVID BIRSGRGDPV RQAARTAEH QDADQRLDR
241 TVAGDVLINL LRPVANARY IFVPAAMLH HDHORALAD GEEAAREPD TDVPRFPFEP
301 FGSKVRPAEF HGGQETYFREF EMVLMDQGR NDREPLWHEP ERPDDGERAP ETIDDDPMVMS
361 HGRGSAADGH RCPQGEGITRI LLKLTSLQQA ATRYTVPPQD LITLDAHVPF RPRSGFVVAR
421 VHRP

FIG. 7L
Bacillus subtilis CYP152A1 (NP_388092) DNA (SEQ ID NO:13)

1  ATGAATGAGC  AGATTCACCA  TGACACAAATG  CTCAAGAACACT  GCTGAAGGAA
61  GGCTATTAT  TTATATACAA  CAGAACAGAG  CAGCTACAGTT  CAGATGTGCT  TGCTGCGCGT
121  TTGGTGAGAA  AAAACTTTAT  TTGCGATTCT  GGGCGCAGGG  GCGCGAGGCT  GTGTTATGAT
181  ACGGCTGAT  TCCGCGGCAA  GAAAGCTTAC  CTTAAAGCTGCC  TCGAGGATGC  GCTGTTTGCT
241  GTGCTTGGCAA  GGTGCGGAG  GCGATGCATCG  ATCGGAGATT  GCTTTTTCTG
301  TGATGTGATC  GACGCCTCCA  TCAGAAGACG  TTGCGTACC  TGAATACAGA  GAGGAGAGAA
361  GGCCACGACTCA  CCAGATGGAG  GAAAAGGATAG  GAGGTTAGTGT  TATITGACGAA  AGCAAGAAAG
421  ATCTCTGAGCC  GCTAGCGCGT  CTTATTGGGCA  GCTGCTCCATG  TGAAGGAAAC  GGAAGTCCAA
481  GAGACGACGCC  ATGACTTCTAT  TGACATGCTCT  GAGCGGCGG  GTGCTGCGGG  ACCCGGCGAT
541  TGAGGAGCAA  GAAGAGCAAG  GCGCGGCGGG  GAAACGGGAG  TTGAGCTAGAT  GATTGAGAT
601  GCTCGTGGCG  GCTGCTCCTG  AAGCACTGCC  GAGCCACGCG  TGGCTGAAAT  GCTTTTTTAC
661  AGACCAACAGG  ATGGAGAGCG  GCTGATTCCC  GCGATGCGGG  CCAATGAGCT  GATTTAGTGA
721  CTGCGGACCTA  TTGCGCTCAG  TTCTCTACCTT  CTGCGTCTCC  CAGCTGGTGG  GCTCTCATGAG
781  CACTCGAAGTC  ATAGGAAATAG  GCTGCGGCTCT  GGAAACAGCC  GGAAGAAGAGA  AATGTGTTG
841  CGAGGACGTCC  GCGATATATT  TCGTGGTCTG  CCGTTTTCGG  GCGCGCTCTGT  CAAAAAGCAT
901  TTTTCTGATGA  ATGACTGTTGA  GTTTAAGAAG  GGCCACAGCGG  TTGCTGCTCTG  TTATATGGA
961  AGCAACCAAGG  ACCCTTCTCT  ATGGGATCAT  GCCGATGAAA  TGGCGCGGGA  ACGATTGCGG
1021  GAGCGGCGAG  AAAATCTGTTT  TGAATGAACT  GCTCAAGCGG  GGGCGGCGCC  GCGAAAGGCG
1081  CACCCTGGTC  CAGGGGAGAA  CATTCAATTG  GAAGTCATGA  AAGGACGCTC  GAGTTTCTCTC
1141  GTCCCATACGA  TGGATACAGA  TGGTGGGGAA  CAACTACCTC  ATTACAGCT  GCGCGAGATAG
1201  CCATGACGGG  CTGAGTAAATG  AGCGGAATCA  GACGAAAGAG  TTAA

FIG. 7M

Bacillus subtilis CYP152A1 (NP_388092) Protein (SEQ ID NO:14)

1  MNEQIPHDKS  LNSNTTLAEE  GYLFPKTRNE  RYNSSDLFQAR  LLGKNPCTMT  GAAAKVFYD
61  TDIFQFQNAL  PKRVKQSLFG  VNAIOQMDGQS  AHIHRMLFL  SLMTFPHQRK  LAELMTESWK
121  AAVTRKWEAD  EVVLPEEAAK  ILCRVACWYA  GVPLKETEVK  ERADDEFDMV  DAFGAVGFRH
181  WGRERARPRF  BEMNNVIMED  ARAGLKLTTTS  GIALHEMAPH  TQEGSGLSIDS  RMAAIELLNV
241  LRTPSWLSTFY  LVPSALALHE  HPKYKELWLS  GNSREREMFV  QEVKRVYYPF  PFLGALVKKD
301  FYWNINCYPPK  GTSVLLDLGY  THDRPRLADH  PDEIRFREVPA  ERBNLMFDDI  FCGSSHABKG
361  HRCPGESGTTI  EVMKASDLDFL  VHQIEZDVPE  QSLHYSLARM  PSLPESGFVM  SGIRRKS

FIG. 7N
FIG. 8
<table>
<thead>
<tr>
<th>Genotype</th>
<th>Gene Product</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>accA</td>
<td>Acetyl-CoA carboxylase, subunit A (carboxyltransferase)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accB</td>
<td>Acetyl-CoA carboxylase, subunit B (BCP: biotin carboxyl protein)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accC</td>
<td>Acetyl-CoA carboxylase, subunit C (biotin carboxylase)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accD</td>
<td>Acetyl-CoA carboxylase, subunit D (carboxyltransferase beta)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accA</td>
<td>Acetyl-CoA carboxylase, subunit A (carboxyltransferase)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accB</td>
<td>Acetyl-CoA carboxylase, subunit B (BCP: biotin carboxyl protein)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accC</td>
<td>Acetyl-CoA carboxylase, subunit C (biotin carboxylase)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
<tr>
<td>accD</td>
<td>Acetyl-CoA carboxylase, subunit D (carboxyltransferase beta)</td>
<td>Increase Malonyl-CoA production</td>
</tr>
</tbody>
</table>

**FIG. 10A**

Accession Numbers are from NCBI, GenBank, Release 159.0 as of April 15, 2007. EC Numbers are from KEGG, Release 42.0 as of April 2007 (plus daily updates up to March, 2008).
<table>
<thead>
<tr>
<th>Gene/Enzyme</th>
<th>Strain/Condition</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Increase Acetyl-CoA production</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td>acfE</td>
<td>E. coli</td>
<td>Over-expression</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>Accession Numbers</td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>fabH</td>
<td>3-oxoacyl-[acyl-carrier-protein] synthase III</td>
<td>AAC74175</td>
</tr>
<tr>
<td>fabI</td>
<td>enoyl-[acyl-carrier-protein] reductase, NADH-dependent</td>
<td>NP_415804</td>
</tr>
<tr>
<td>fabR</td>
<td>Transcriptional Repressor</td>
<td>NP_418398</td>
</tr>
<tr>
<td>fabZ</td>
<td>(3R)-hydroxymyristol acyl carrier protein dehydratase</td>
<td>NP_414722</td>
</tr>
<tr>
<td>fabE</td>
<td>acyl-CoA dehydrogenase</td>
<td>AAC73325</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione synthase</td>
<td>P04425</td>
</tr>
<tr>
<td>gspA</td>
<td>biosynthetic sn-glycerol 3-phosphate dehydrogenase</td>
<td>AAC76632, NP_41805</td>
</tr>
<tr>
<td>lhdA</td>
<td>lactate dehydrogenase</td>
<td>AAC74462, NP_415898</td>
</tr>
<tr>
<td>Lipase</td>
<td>Triglyceride Lipase</td>
<td>CAA89087, CAA98876</td>
</tr>
<tr>
<td></td>
<td>Malonyl-CoA decarboxylase</td>
<td>AAA26500</td>
</tr>
<tr>
<td>penD</td>
<td>aspartate f- decarboxylase</td>
<td>BAB96708</td>
</tr>
<tr>
<td>Gene</td>
<td>a.k.a.</td>
<td>Function</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>pank</td>
<td>a.k.a.</td>
<td>pantothenate kinase</td>
</tr>
<tr>
<td>pank</td>
<td>a.k.a.</td>
<td>pantothenate kinase</td>
</tr>
<tr>
<td>pdh</td>
<td></td>
<td>Pyruvate dehydrogenase</td>
</tr>
<tr>
<td>pflB</td>
<td></td>
<td>formate acetyltransferase (pyruvate formate lyase)</td>
</tr>
<tr>
<td>plsB</td>
<td></td>
<td>acyltransferase</td>
</tr>
<tr>
<td>poxB</td>
<td></td>
<td>pyruvate oxidase</td>
</tr>
<tr>
<td>pta</td>
<td></td>
<td>phosphotransacetylase</td>
</tr>
<tr>
<td>udhA</td>
<td></td>
<td>pyridine nucleotide transhydrogenase</td>
</tr>
<tr>
<td>fadB</td>
<td></td>
<td>fused 3-hydroxybutyryl-CoA epimerase/delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase/enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>3-hydroxyacyl-CoA dehydrogenase; K01692 enoyl-CoA hydratase; K01782 3-hydroxybutyryl-CoA epimerase</td>
<td>AAC75401</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>fadA</td>
<td>3-ketoacyl-CoA thiolase</td>
<td>BAE77458</td>
</tr>
<tr>
<td>fadl</td>
<td>beta-ketoacyl-CoA thiolase</td>
<td>AAC75402</td>
</tr>
<tr>
<td>YdiO</td>
<td>acyl-CoA dehydrogenase</td>
<td>YP_852786</td>
</tr>
</tbody>
</table>

2. Structure Control

2A. Chain Length Control

<table>
<thead>
<tr>
<th></th>
<th>thioesterase</th>
<th>POADA1</th>
<th>3.1.2.11, 3.1.1.5</th>
<th>Delete and/or express</th>
<th>C18 Chain Length</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>tesA</td>
<td>thioesterase</td>
<td>AAC73596, NP_415027</td>
<td>3.1.2.11, 3.1.1.5</td>
<td>express or overexpress</td>
<td>C18:1</td>
<td>E. coli</td>
</tr>
<tr>
<td>tesA without leader sequence</td>
<td>thioesterase</td>
<td>POADA1</td>
<td>3.1.2.11, 3.1.1.5</td>
<td>Express and/or overexpress mutation L109P</td>
<td>&lt;C18 Chain Length</td>
<td>E. coli</td>
</tr>
<tr>
<td>fatB1 (umbellularia)</td>
<td>thioesterase</td>
<td>Q41635</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C12:0</td>
<td>Umbellularia californica</td>
</tr>
<tr>
<td>fatB2 (umbellularia)</td>
<td>thioesterase</td>
<td>AAC49269</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C8:0 - C10:0</td>
<td>Cuphea hookeriana</td>
</tr>
<tr>
<td>fatB3</td>
<td>thioesterase</td>
<td>AAC72881</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C14:0 - C16:0</td>
<td>Cuphea hookeriana</td>
</tr>
<tr>
<td>Protein</td>
<td>Enzyme</td>
<td>Accession No.</td>
<td>Date</td>
<td>Expression</td>
<td>Product</td>
<td>Source</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>---------------</td>
<td>------</td>
<td>------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>fatB (Cinnamomum)</td>
<td>thioesterase</td>
<td>Q39473</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C14:0</td>
<td>Cinnamomum camphora</td>
</tr>
<tr>
<td>fatB(M141T)</td>
<td>thioesterase</td>
<td>CAA85388</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C16:1</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>fatA1 (Helianthus)</td>
<td>thioesterase</td>
<td>AAL79361</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C18:1</td>
<td>Helianthus annuus</td>
</tr>
<tr>
<td>attata (ARABIDOPSIS FATA ACYL-ACP THIOESTERASE)</td>
<td>thioesterase</td>
<td>NP_189147, NP_193041</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C18:1</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>fatA</td>
<td>thioesterase</td>
<td>CAC39106</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C18:1</td>
<td>Brassica juncea</td>
</tr>
<tr>
<td>fatA (cuphea)</td>
<td>thioesterase</td>
<td>AAC72883</td>
<td>3.1.2.14</td>
<td>express or overexpress</td>
<td>C18:1</td>
<td>Cuphea hookeriana</td>
</tr>
</tbody>
</table>

2B. Branching Control

- attenuate FabH
- express FabH from *S. glaucescens* or *S. coelicolor* and knock out endogenous FabH
- express FabH from *B. subtilis* and knock out endogenous FabH

- increase branched chain fatty acid derivatives
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>EC</th>
<th>Expression</th>
<th>Source</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>bkdA1</td>
<td>branched-chain-3-ketoacyl-acyl-carrier protein</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>NP_628006</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>bkdB1</td>
<td>branched-chain-3-ketoacyl-acyl-carrier protein</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>NP_628005</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>bkdC1</td>
<td>branched-chain-3-ketoacyl-acyl-carrier protein</td>
<td>2.3.1.168</td>
<td>express or Over-Express</td>
<td>NP_628004</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>bkdA2</td>
<td>branched-chain-3-ketoacyl-acyl-carrier protein</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>NP_733618</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Accession</td>
<td>EC Class</td>
<td>Expression</td>
<td>Products</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>bkdB2</td>
<td>branched-chain a-ketoacid dehydrogenase b-subunit (E1b)</td>
<td>NP_628019</td>
<td>EC 1.2.4.4</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdC2</td>
<td>dihydrolipoyl transacetylase (E2)</td>
<td>NP_628018</td>
<td>EC 2.3.1.168</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdA</td>
<td>branched-chain a-ketoacid dehydrogenase a-subunit (E1a)</td>
<td>BAC72074</td>
<td>EC 1.2.4.4</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdB</td>
<td>branched-chain a-ketoacid dehydrogenase b-subunit (E1b)</td>
<td>BAC72075</td>
<td>EC 1.2.4.4</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdC</td>
<td>dihydrolipoyl transacetylase (E2)</td>
<td>BAC72076</td>
<td>EC 2.3.1.168</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdF</td>
<td>branched-chain a-ketoacid dehydrogenase a-subunit (E1a)</td>
<td>BAC72078</td>
<td>EC 1.2.4.4</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdG</td>
<td>branched-chain a-ketoacid dehydrogenase b-subunit (E1b)</td>
<td>BAC72089</td>
<td>EC 1.2.4.4</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td>bkdH</td>
<td>dihydrolipoyl transacetylase (E2)</td>
<td>BAC72090</td>
<td>EC 2.3.1.168</td>
<td>express or Over-Express</td>
<td>make branched-chain acyl-CoA precursors</td>
</tr>
<tr>
<td></td>
<td>protein</td>
<td>accession</td>
<td>EC</td>
<td>expression</td>
<td>source</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>-----------</td>
<td>------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>bkdAA</td>
<td>branched-chain a-ketoacid dehydrogenase a-subunit (E1a)</td>
<td>NP_390285</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>bkdAB</td>
<td>branched-chain a-ketoacid dehydrogenase b-subunit (E1b)</td>
<td>NP_390284</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>bkdB</td>
<td>dihydrolipoyl transacetylase (E2)</td>
<td>NP_390283</td>
<td>2.3.1.168</td>
<td>express or Over-Express</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>bkdA1</td>
<td>branched-chain a-ketoacid dehydrogenase a-subunit (E1a)</td>
<td>AAA65614</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>bkdA2</td>
<td>branched-chain a-ketoacid dehydrogenase b-subunit (E1b)</td>
<td>AAA65615</td>
<td>1.2.4.4</td>
<td>express or Over-Express</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>bkdC</td>
<td>dihydrolipoyl transacetylase (E2)</td>
<td>AAA65617</td>
<td>2.3.1.168</td>
<td>express or Over-Express</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Ipd</td>
<td>dihydrolipoamide dehydrogenase (E3)</td>
<td>NP_414658</td>
<td>1.8.1.4</td>
<td>express or Over-Express</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>lME</td>
<td>branched-chain amino acid aminotransferase</td>
<td>YP_026247</td>
<td>2.6.1.42</td>
<td>express or Over-Express</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>lME</td>
<td>branched-chain amino acid aminotransferase</td>
<td>AAF34406</td>
<td>2.6.1.42</td>
<td>express or Over-Express</td>
<td>Lactococcus lactis</td>
</tr>
</tbody>
</table>
### FIG. 10J

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lvmE</td>
<td>branched-chain amino acid aminotransferase</td>
<td>NP_745648</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>lvmE</td>
<td>branched-chain amino acid aminotransferase</td>
<td>NP_629657</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>cor</td>
<td>crotonyl-CoA reductase</td>
<td>NP_630556</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>cor</td>
<td>crotonyl-CoA reductase</td>
<td>AAD53915</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>IcmA, isobutyryl-CoA mutase, subunit A</td>
<td>isobutyryl-CoA mutase, subunit A</td>
<td>NP_629554</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>IcmA, isobutyryl-CoA mutase, subunit A</td>
<td>isobutyryl-CoA mutase, subunit A</td>
<td>AAC08713</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>IcmB, isobutyryl-CoA mutase, subunit B</td>
<td>isobutyryl-CoA mutase, subunit B</td>
<td>NP_630904</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>IcmB, isobutyryl-CoA mutase, subunit B</td>
<td>isobutyryl-CoA mutase, subunit B</td>
<td>CAB59633</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>FabH, ACPs and fabF genes with specificity for branched chain acyl-CoAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lvmE</td>
<td>branched-chain amino acid aminotransferase</td>
<td>CAC12788</td>
<td>EC2.6.1.4</td>
</tr>
<tr>
<td></td>
<td>beta-ketoacyl-ACP synthase III</td>
<td>NP_626634</td>
<td>2.3.1.180</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>FabH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACP</td>
<td>acyl-carrier protein</td>
<td>NP_626635</td>
<td>NONE</td>
</tr>
<tr>
<td>FabF</td>
<td>beta-ketoacyl-ACP synthase II</td>
<td>NP_626636</td>
<td>2.3.1.179</td>
</tr>
<tr>
<td>FabH3</td>
<td>beta-ketoacyl-ACP synthase III</td>
<td>NP_823466</td>
<td>2.3.1.180</td>
</tr>
<tr>
<td>FabC3 (ACP)</td>
<td>acyl-carrier protein</td>
<td>NP_823467</td>
<td>NONE</td>
</tr>
<tr>
<td>FabF</td>
<td>beta-ketoacyl-ACP synthase II</td>
<td>NP_823468</td>
<td>2.3.1.179</td>
</tr>
<tr>
<td>Figure</td>
<td>Enzyme Name</td>
<td>Reference</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Fig. 10L</td>
<td>FabH A</td>
<td>NC_039015</td>
<td>beta-ketoacyl-ACP synthase III (encoded by fabH)</td>
</tr>
<tr>
<td></td>
<td>FabH B</td>
<td>NC_038896</td>
<td>beta-ketoacyl-ACP synthase III (encoded by fabH)</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>NC_038897</td>
<td>beta-ketoacyl-ACP synthase II (encoded by fabH)</td>
</tr>
<tr>
<td></td>
<td>FabF</td>
<td>NC_014462</td>
<td>beta-ketoacyl-ACP synthase II (encoded by fabH)</td>
</tr>
<tr>
<td></td>
<td>SmallDRAFT_081</td>
<td>NC_014462</td>
<td>beta-ketoacyl-ACP synthase II (encoded by fabH)</td>
</tr>
<tr>
<td></td>
<td>SmallDRAFT_082</td>
<td>NC_014462</td>
<td>beta-ketoacyl-ACP synthase II (encoded by fabH)</td>
</tr>
</tbody>
</table>

**Legend:**
- *Bacillus subtilis*
- *Sterotrophomonas maltophilia*
<table>
<thead>
<tr>
<th></th>
<th><strong>FABH</strong></th>
<th><strong>ACP</strong></th>
<th><strong>FABF</strong></th>
<th><strong>FABH</strong></th>
<th><strong>FABF</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YP_123672</td>
<td>YP_123675</td>
<td>YP_123676</td>
<td>NIP_415609</td>
<td>NIP_415613</td>
</tr>
<tr>
<td></td>
<td>2.3.1.180</td>
<td>NONE</td>
<td>2.3.1.179</td>
<td>2.3.1.180</td>
<td>2.3.1.179</td>
</tr>
<tr>
<td></td>
<td>express or Over-Express</td>
<td>express or Over-Express</td>
<td>express or Over-Express</td>
<td>delete or reduce</td>
<td>delete or reduce</td>
</tr>
<tr>
<td></td>
<td>initiation of branched-chain fatty acid biosynthesis</td>
<td>initiation and elongation of branched-chain fatty acid biosynthesis</td>
<td>elongation of branched-chain fatty acid biosynthesis</td>
<td>initiation of branched-chain fatty acid biosynthesis</td>
<td>elongation of branched-chain fatty acid biosynthesis</td>
</tr>
<tr>
<td></td>
<td><strong>Legionella pneumophila</strong></td>
<td><strong>Legionella pneumophila</strong></td>
<td><strong>Legionella pneumophila</strong></td>
<td><strong>Escherichia coli</strong></td>
<td><strong>Escherichia coli</strong></td>
</tr>
</tbody>
</table>

**FIG. 10M**

To Produce Cyclic Fatty Acids

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AnsJ</td>
<td>dehydratase (putative)</td>
<td>not available</td>
<td>not available</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>AnsK</td>
<td>CoA ligase (putative)</td>
<td>not available</td>
<td>not available</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td>AnsL</td>
<td>dehydrogenase (putative)</td>
<td>not available</td>
<td>not available</td>
<td>express or Over-Express</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cyclohexylcarboxylic-CoA biosynthesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cyclohexylcarboxylic-CoA biosynthesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cyclohexylcarboxylic-CoA biosynthesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Enzyme</td>
<td>Function</td>
<td>Accession No.</td>
<td>EC Number</td>
<td>Expression</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------</td>
<td>---------------</td>
<td>-----------</td>
<td>--------------</td>
</tr>
<tr>
<td>ChcA</td>
<td>enoyl-CoA reductase</td>
<td>U72144</td>
<td>EC 1.3.1.34</td>
<td>express or</td>
</tr>
<tr>
<td>AnsM</td>
<td>oxidoreductase (putative)</td>
<td>not available</td>
<td>not available</td>
<td>Over-Express</td>
</tr>
<tr>
<td>PmlJ</td>
<td>dehydratase (putative)</td>
<td>AAA84158</td>
<td>not available</td>
<td>express or</td>
</tr>
<tr>
<td>PmlK</td>
<td>CoA ligase (putative)</td>
<td>AAA84158</td>
<td>not available</td>
<td>express or</td>
</tr>
<tr>
<td>PmlL</td>
<td>dehydrogenase (putative)</td>
<td>AAA84159</td>
<td>not available</td>
<td>express or</td>
</tr>
<tr>
<td>ChcA</td>
<td>enoyl-CoA reductase</td>
<td>AAA84160</td>
<td>EC 1.3.1.34</td>
<td>express or</td>
</tr>
<tr>
<td>PmlM</td>
<td>oxidoreductase (putative)</td>
<td>AAA84161</td>
<td>not available</td>
<td>express or</td>
</tr>
<tr>
<td>ChcB</td>
<td>enoyl-CoA isomerase</td>
<td>AF268489</td>
<td>not available</td>
<td>express or</td>
</tr>
<tr>
<td>ChcB/CatD</td>
<td>enoyl-CoA isomerase</td>
<td>NP_629292</td>
<td>4.2.1.-</td>
<td>express or</td>
</tr>
<tr>
<td>ChcB/CatD</td>
<td>enoyl-CoA isomerase</td>
<td>NP_824296</td>
<td>4.2.1.-</td>
<td>express or</td>
</tr>
<tr>
<td></td>
<td>2C. Saturation Level Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Strain</td>
<td>Suppressor of FabA</td>
<td>E. coli</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AAC7592, AAC4390</td>
<td>NONE</td>
<td>OVEREXPRESS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>also see fabA in section 1</td>
<td>also see section 2A: items with ( \delta = 0 ) are unsaturated (no double bonds) and with ( \delta = 1 ) are saturated (1 double bond)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>fabB</td>
<td>3-acyl-CoA:acyl-CoA hydrolase I</td>
<td>E. coli</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fabK</td>
<td>trans-2-acyl-ACP reductase II</td>
<td>AAP8073</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fabL</td>
<td>enoyl-[acyl carrier protein] reductase</td>
<td>AAP8621</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table contains information on strain names, suppressors, and enzymes related to fatty acid metabolism, specifically focusing on saturation levels and their corresponding functional expressions.
<table>
<thead>
<tr>
<th>Transposon</th>
<th>Function</th>
<th>Gene</th>
<th>Protein</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>fadB</td>
<td>ABC transport protein</td>
<td>CEP5</td>
<td>Arabidopsis thaliana</td>
<td>Export products</td>
</tr>
<tr>
<td>fabM</td>
<td>ABC transport protein</td>
<td>At2g46740, At5g1460</td>
<td>Arabidopsis thaliana</td>
<td>Export products</td>
</tr>
<tr>
<td>DAL05901</td>
<td>4,2,1,17</td>
<td>AAT17328</td>
<td>Solanum tuberosum</td>
<td>Export products</td>
</tr>
<tr>
<td>trans-2,3-decanoyl-ACP isomerase</td>
<td>modulate unsaturated fatty acid production</td>
<td>Arabinobiosis thaliana</td>
<td>Arabidopsis thaliana</td>
<td>Export products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rhodococcus erythropolis</td>
<td>Arabidopsis thaliana</td>
<td>Export products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B. subtilis</td>
<td>Protocallimyxa arnvebrillia</td>
<td>Export products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C. albicans</td>
<td>UME25</td>
<td>Export products</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F. tularensis subsp. novicida</td>
<td>C. albicans</td>
<td>Export products</td>
</tr>
</tbody>
</table>

**FIG. 10P**
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Accession</th>
<th>Expression</th>
<th>Function</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcrE</td>
<td>Transmembrane protein affects septum formation and cell membrane permeability</td>
<td>YP_312213</td>
<td>NONE</td>
<td>express</td>
<td><em>Shigella sonnei</em> S5046</td>
</tr>
<tr>
<td>AcrF</td>
<td>Acriflavine resistance protein F</td>
<td>P24181</td>
<td>NONE</td>
<td>express</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>tll1618</td>
<td>Multidrug efflux transporter</td>
<td>NP_682408.1</td>
<td>NONE</td>
<td>express</td>
<td><em>Thermosynechococcus elongatus</em> BP-1</td>
</tr>
<tr>
<td>tll1619</td>
<td>Multidrug efflux transporter</td>
<td>NP_682409.1</td>
<td>NONE</td>
<td>express</td>
<td><em>Thermosynechococcus elongatus</em> BP-1</td>
</tr>
<tr>
<td>tll0139</td>
<td>Multidrug efflux transporter</td>
<td>NP_680930.1</td>
<td>NONE</td>
<td>express</td>
<td><em>Thermosynechococcus elongatus</em> BP-1</td>
</tr>
</tbody>
</table>

4. Fermentation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Accession</th>
<th>Expression</th>
<th>Function</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>umuD</td>
<td>DNA polymerase V, subunit</td>
<td>YP_310132</td>
<td>3.4.21.-</td>
<td>Over-express</td>
<td><em>Shigella sonnei</em> S5046</td>
</tr>
<tr>
<td>umuC</td>
<td>DNA polymerase V, subunit</td>
<td>ABC42261</td>
<td>2.7.7.7</td>
<td>Over-express</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>NADH:NADPH transhydrogenase (alpha and beta subunits) (pntA, pntB)</td>
<td>P07001, P04870</td>
<td>1.6.1.2</td>
<td>express</td>
<td><em>Shigella flexneri</em></td>
<td></td>
</tr>
</tbody>
</table>
FIG. 11
FIG. 13A

FIG. 13B

FIG. 13C

1-heptadecadiene

1-heptadecene

ORF2459 + C18 feeding
FIG. 15A

TIC: emiQ27.D\data.ms

1-heptadecane

1-octadecane (internal standard)

FIG. 15B

TIC: emiQ36.D\data.ms

FIG. 15C

TIC: emiQ43.D\data.ms

pHT01_orf980 (Jeogalicoccus)

pHT01_orf2459 (C. efficiens)

pHT01
FIG. 17A

1-pentadecene

orfKRH2157

FIG. 17B

No fatty acid

FIG. 17C

No protein
FIG. 18A

Abundance

Time →

1-pentadecene

orf_Mpop129

FIG. 18B

Abundance

Time →

No fatty acid

FIG. 18C

Abundance

Time →

No protein
FIG. 20

Ratio of Decarboxylation to Hydroxylation

FIG. 21
<table>
<thead>
<tr>
<th>Organism</th>
<th>GeneBank Acc. #</th>
<th>Pos.</th>
<th>%ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeotgalicoccus sp. ATCC8456</td>
<td>orf880</td>
<td>His</td>
<td>100</td>
</tr>
<tr>
<td>Arthrobacter sp. FB24</td>
<td>ZP_00413206</td>
<td>Gln</td>
<td>32.7</td>
</tr>
<tr>
<td>Aurantimonas sp. SI85-9A1</td>
<td>ZP_01227146</td>
<td>Gln</td>
<td>34.7</td>
</tr>
<tr>
<td>Bacillus amyloidefaciens FZB42</td>
<td>ABS73471</td>
<td>Gln</td>
<td>40.8</td>
</tr>
<tr>
<td>Bacillus claussenii KSM-K16</td>
<td>YP_176535</td>
<td>Gln</td>
<td>37.4</td>
</tr>
<tr>
<td>Bacillus cohaillienis m4-4</td>
<td>ZP_03227611</td>
<td>Gln</td>
<td>40.2</td>
</tr>
<tr>
<td>Bacillus coagulans 36D1</td>
<td>EAY45396</td>
<td>Gln</td>
<td>37.1</td>
</tr>
<tr>
<td>Bacillus halodurans C-125</td>
<td>NP_242623</td>
<td>Gln</td>
<td>35.9</td>
</tr>
<tr>
<td>Bacillus subtilis str. 168</td>
<td>NP_388092</td>
<td>Gln</td>
<td>41.2</td>
</tr>
<tr>
<td>Bacillus sp. NRRLB-14911</td>
<td>ZP_0116817</td>
<td>Gln</td>
<td>39.6</td>
</tr>
<tr>
<td>Brevundimonas sp. BAL3</td>
<td>EDX80366</td>
<td>Gln</td>
<td>32.4</td>
</tr>
<tr>
<td>Clostridium acetobutylicum ATCC 824</td>
<td>NP_349922</td>
<td>Gln</td>
<td>39.1</td>
</tr>
<tr>
<td>Clostridium botulinum F str. Langeland</td>
<td>ABS41685</td>
<td>Gln</td>
<td>40.2</td>
</tr>
<tr>
<td>Clostridium botulinum NCTC 2916</td>
<td>ZP_02613119</td>
<td>Gln</td>
<td>40.7</td>
</tr>
<tr>
<td>Clostridium botulinum A str. ATCC 19397</td>
<td>ABS33032</td>
<td>Gln</td>
<td>40.0</td>
</tr>
<tr>
<td>Clostridium botulinum Bf</td>
<td>ZP_02617555</td>
<td>Gln</td>
<td>40.7</td>
</tr>
<tr>
<td>Clostridium botulinum A3 str. Loch Maree</td>
<td>YP_001786433</td>
<td>Gln</td>
<td>40.5</td>
</tr>
<tr>
<td>Clostridium botulinum B1 str. Okra</td>
<td>YP_00178672</td>
<td>Gln</td>
<td>40.0</td>
</tr>
<tr>
<td>Clostridium botulinum A str. ATCC 3502</td>
<td>YP_001253569</td>
<td>Gln</td>
<td>40.0</td>
</tr>
<tr>
<td>Clostridium phytofermentans ISDg</td>
<td>ZP_01354055</td>
<td>Gln</td>
<td>37.1</td>
</tr>
<tr>
<td>Clostridium sporogenes ATCC 15579</td>
<td>ZP_02994641</td>
<td>Gln</td>
<td>40.0</td>
</tr>
<tr>
<td>Corynebacterium efficiens YS-314</td>
<td>NP_739069</td>
<td>His</td>
<td>28.8</td>
</tr>
<tr>
<td>Cyanobacter sp. CCY0110</td>
<td>EAZ92180</td>
<td>Gln</td>
<td>36.2</td>
</tr>
<tr>
<td>Enterococcus faecium DO</td>
<td>ZP_00604550</td>
<td>Gln</td>
<td>28.4</td>
</tr>
<tr>
<td>Exiguobacterium sp. AT1b</td>
<td>ZP_02991092</td>
<td>Gln</td>
<td>34.5</td>
</tr>
<tr>
<td>Kocuria rhizophila DC2201</td>
<td>YP_001856010</td>
<td>His</td>
<td>29.8</td>
</tr>
<tr>
<td>Mespilirhodobacterium sp. BNC1</td>
<td>YP_665837</td>
<td>Gln</td>
<td>35.2</td>
</tr>
<tr>
<td>Mespilirhodobacterium sp. BNC1</td>
<td>YP_676050</td>
<td>Gln</td>
<td>29.9</td>
</tr>
<tr>
<td>Methylobacillus flagellatus KT</td>
<td>YP_545551</td>
<td>Gln</td>
<td>34.4</td>
</tr>
<tr>
<td>Methylobacterium populorum BJ001</td>
<td>EDO73824</td>
<td>Met</td>
<td>31.3</td>
</tr>
<tr>
<td>Methylobacterium sp. 4-4</td>
<td>EDK95966</td>
<td>Met</td>
<td>35.2</td>
</tr>
<tr>
<td>Oceanicola baltensis HTCC2597</td>
<td>ZP_01000316</td>
<td>Gln</td>
<td>30.8</td>
</tr>
<tr>
<td>Oceanobacillus ilevensis HTE831</td>
<td>NP_694023</td>
<td>Phe</td>
<td>30.9</td>
</tr>
<tr>
<td>Pseudomonas stutzeri A1501</td>
<td>ABB80313</td>
<td>Gln</td>
<td>32.9</td>
</tr>
<tr>
<td>Planctomyces maris DSM 8797</td>
<td>EDL56759</td>
<td>Gln</td>
<td>32.0</td>
</tr>
<tr>
<td>Rhodobacter sphaeroides 2.4.1</td>
<td>YP_362437</td>
<td>Gln</td>
<td>30.9</td>
</tr>
<tr>
<td>Rhodobacter sphaeroides ATCC 17025</td>
<td>ZP_00915817</td>
<td>Gln</td>
<td>30.6</td>
</tr>
<tr>
<td>Rhodobacter sphaeroides ATCC 17029</td>
<td>ZP_00916496</td>
<td>Gln</td>
<td>31.1</td>
</tr>
<tr>
<td>Sphingomonas paucimobilis</td>
<td>BAA22987</td>
<td>Gln</td>
<td>36.0</td>
</tr>
<tr>
<td>Sphingopyxis alaskensis RB2256</td>
<td>YP_617595</td>
<td>Gln</td>
<td>35.2</td>
</tr>
<tr>
<td>Streptomyces albus</td>
<td>ABS90483</td>
<td>His</td>
<td>34.6</td>
</tr>
<tr>
<td>Streptomyces pristinae spiralis ATCC 25486</td>
<td>YP_002196443</td>
<td>His</td>
<td>33.2</td>
</tr>
<tr>
<td>Streptomyces sp. Mg1</td>
<td>ZP_03175040</td>
<td>His</td>
<td>34.9</td>
</tr>
<tr>
<td>Sulfobacter sp. NAS-14.1</td>
<td>ZP_00964321</td>
<td>Gln</td>
<td>34.1</td>
</tr>
<tr>
<td>Xanthomonas campestris str. 8004</td>
<td>YP_243267</td>
<td>Gln</td>
<td>27.4</td>
</tr>
</tbody>
</table>

1: in relation to Jeotgalicoccus orf880

**FIG. 22**
METHODS AND COMPOSITIONS FOR PRODUCING OLEFINs

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Pat. No. 8,597,922, filed Apr. 20, 2012, which is a divisional of U.S. Pat. No. 8,183,028, filed Apr. 14, 2011, which is the National Stage of International Application No. PCT/US08/14029, filed Dec. 22, 2008, which claims the benefit of U.S. Provisional Application No. 61/092,278, filed on Aug. 27, 2008; U.S. Provisional Application No. 61/051,886, filed on May 9, 2008; and U.S. Provisional Application No. 61/016, 183, filed on Dec. 21, 2007, all of which are incorporated herein by reference in their entirety.

INCORPORATION-BY-REFERENCE OF MATERIAL SUBMITTED ELECTRONICALLY

[0002] Incorporated by reference in its entirety herein is a computer-readable nucleotide/amino acid sequence listing submitted in International Application No. PCT/US08/14029, filed Dec. 22, 2008, and identified as follows: One 40,960 Byte ASCII (Text) file named "L300011 PCT ST25_Sequence List_05.12.09" created on Mar. 11, 2009.

BACKGROUND OF THE INVENTION

[0003] Petroleum is a limited, natural resource found in the Earth in liquid, gaseous, or solid forms. Petroleum is primarily composed of hydrocarbons, which are comprised mainly of carbon and hydrogen. It also contains significant amounts of other elements, such as, nitrogen, oxygen, or sulfur, in different forms.

[0004] Petroleum is a valuable resource, but petroleum products are developed at considerable costs, both financial and environmental. First, sources of petroleum must be discovered. Petroleum exploration is an expensive and risky venture. The cost of exploring deep water wells can exceed $100 million. Moreover, there is no guarantee that these wells will contain petroleum. It is estimated that only 40% of drilled wells lead to productive wells generating commercial hydrocarbons. In addition to the economic cost, petroleum exploration carries a high environmental cost. For example, offshore exploration disturbs the surrounding marine environments.

[0005] After a productive well is discovered, the petroleum must be extracted from the Earth at great expense. During primary recovery, the natural pressure underground is sufficient to extract about 20% of the petroleum in the well. As this natural pressure falls, secondary recovery methods are employed, if economical. Generally, secondary recovery involves increasing the well’s pressure by, for example, water injection, natural gas injection, or gas lift. Using secondary recovery methods, an additional 5% to 15% of petroleum is recovered. Once secondary recovery methods are exhausted, tertiary recovery methods can be used, if economical. Tertiary methods involve reducing the viscosity of the petroleum to make it easier to extract. Using tertiary recovery methods, an additional 5% to 15% of petroleum is recovered. Hence, even under the best circumstances, only 50% of the petroleum in a well can be extracted. Petroleum extraction also carries an environmental cost. For example, petroleum extraction can result in large see pages of petroleum rising to the surface. Moreover, offshore drilling involves dredging the seabed which disrupts or destroys the surrounding marine environment. Since petroleum deposits are not found uniformly throughout the Earth, petroleum must be transported over great distances from petroleum producing regions to petroleum consuming regions. In addition to the shipping costs, there is also the environmental risk of devastating oil spills.

[0006] In its natural form, crude petroleum extracted from the Earth has few commercial uses. It is a mixture of hydrocarbons (e.g., paraffins or alkanes), olefins (or alkynes), aromatics, and oils (or cycloalkanes), aliphatic compounds, aromatic compounds, etc.) of varying length and complexity. In addition, crude petroleum contains other organic compounds (e.g., organic compounds containing nitrogen, oxygen, sulfur, etc.) and impurities (e.g., sulfur, salt, acid, metals, etc.).

[0007] Hence, crude petroleum must be refined and purified before it can be used commercially. Due to its high energy density and its easy transportability, most petroleum is refined into fuels, such as transportation fuels (e.g., gasoline, diesel, aviation fuel, etc.), heating oil, liquefied petroleum gas, etc.

[0008] Crude petroleum is also a primary source of raw materials for producing petrochemicals. The two main classes of raw materials derived from petroleum are short chain olefins (e.g., ethylene and propylene) and aromatics (e.g., benzene and xylene isomers). These raw materials are derived from longer chain hydrocarbons in crude petroleum by cracking it at considerable expense using a variety of methods, such as catalytic cracking, steam cracking, or catalytic reforming. These raw materials are used to make petrochemicals, which cannot be directly refined from crude petroleum, such as monomers, solvents, detergents, or adhesives.

[0009] One example of a raw material derived from crude petroleum is ethylene. Ethylene is used to produce petrochemicals such as, polyethylene, ethane, ethylene oxide, ethylene glycol, polyester, glycol ether, ethoxyethane, vinyl acetate, 1,2-dichloroethane, trichloroethylene, tetrachloroethylene, vinyl chloride, and polyvinyl chloride. An additional example of a raw material is propylene, which is used to produce isopropyl alcohol, acrylonitrile, polypropylene, propylene oxide, propylene glycol, glycol ethers, butylene, isobutylene, 1,3-butadiene, synthetic elastomers, polyolefins, alpha-olefins, fatty alcohols, acrylic acid, acryl polymers, allyl chloride, epichlorhydrin, and epoxy resin.

[0010] These petrochemicals can then be used to make specialty chemicals, such as plastics, resins, fibers, elastomers, pharmaceuticals, lubricants, or gels. Particular specialty chemicals which can be produced from petrochemical raw materials are: fatty acids, hydrocarbons (e.g., long chain, branched chain, saturated, unsaturated, etc.), fatty alcohols, esters, fatty aldehydes, ketones, lubricants, etc.

[0011] Specialty chemicals have many commercial uses. Hydrocarbons have many commercial uses. For example, shorter chain alkanes are used as fuels. Methane and ethane are the main constituents of natural gas. Longer chain alkanes (e.g., from five to sixteen carbons) are used as transportation fuels (e.g., gasoline, diesel, or aviation fuel). Alkanes having more than sixteen carbon atoms are important components of fuel oils and lubricating oils. Even longer alkanes, which are solid at room temperature, can be used, for example, as a paraffin wax. Alkanes that contain approximately thirty-five carbons are found in bitumen,
which is used for road surfacing. In addition, longer chain alkanes can be cracked to produce commercially useful shorter chain hydrocarbons.

[0012] Like short chain alkanes, short chain olefins, or alkenes, are used in transportation fuels. Longer chain olefins are used in plastics, lubricants, and synthetic lubricants. In addition, olefins are used as a feedstock to produce alcohols, esters, plasticizers, surfactants, tertiary amines, enhanced oil recovery agents, fatty acids, thiols, alkenylsuccinic anhydrides, polymers, textile solvents, adhesives, epoxides, chlorinated alkanes, chlorinated olefins, waxes, fuel additives, and drag flow reducers. In addition, long chain olefins can be cracked to produce fuels.

[0013] Olefins have traditionally been produced from petroleum sources through the process of steam or catalytic cracking. Because of the limited availability and high cost of petroleum sources, the cost of producing olefins from such petroleum sources has been steadily increasing.

[0014] Olefins are the largest volume of chemical intermediates produced in the chemical industry, with global annual production previously estimated at over 300 billion lbs per year. Olefins are produced almost exclusively from ethane or other light alkanes (naphtha) in a process called cracking. This process involves heating the ethane or other light olefins to approximately 750-1000°C in a cracker. It has been estimated that 30% of all pollution from chemical plants comes from cracking owing to emissions and unburned hydrocarbons in the flame required to heat the cracker. Approximately 10% of petroleum is consumed in the production of olefins and related chemicals.

[0015] In addition, crude petroleum is a source of lubricants. Lubricants derived petroleum are typically composed of olefins, particularly polyolefins and terminal olefins. Lubricants can either be refined from crude petroleum or manufactured using raw materials refined from crude petroleum.

[0016] Obtaining these specialty chemicals from crude petroleum requires a significant financial investment as well as a great deal of energy. It is also an inefficient process because frequently the long chain hydrocarbons in crude petroleum are cracked to produce smaller monomers. These monomers are then used as the raw material to manufacture the more complex specialty chemicals.

[0017] In addition to the problems with exploring, extracting, transporting, and refining petroleum, petroleum is a limited and dwindling resource. One estimate of world petroleum consumption is 30 billion barrels per year. By some estimates, it is predicted that at current production levels, the world’s petroleum reserves could be depleted before the year 2050.

[0018] Finally, the burning of petroleum based fuels releases greenhouse gases (e.g., carbon dioxide) and other forms of air pollution (e.g., carbon monoxide, sulfur dioxide, etc.). As the world’s demand for fuel increases, the emission of greenhouse gases and other forms of air pollution also increases. The accumulation of greenhouse gases in the atmosphere leads to an increase in global warming. Hence, in addition to damaging the environment locally (e.g., oil spills, dredging of marine environments, etc.), burning petroleum also damages the environment globally.

[0019] Due to the inherent challenges posed by petroleum, there is a need for a renewable petroleum source which does not need to be explored, extracted, transported over long distances, or substantially refined like petroleum. There is also a need for a renewable petroleum source which can be produced economically without creating the type of environmental damage produced by the petroleum industry and the burning of petroleum based fuels. For similar reasons, there is also a need for a renewable source of chemicals which are typically derived from petroleum.

[0020] One method of producing renewable petroleum is by engineering microorganisms to produce renewable petroleum products. Some microorganisms have a natural ability to produce chemicals. For example, yeast has been used for centuries to produce ethanol (e.g., beer, wine, etc.). In recent years, through the development of advanced biotechnologies, it is possible to metabolically engineer an organism to produce biochemicals that were never previously produced. Chemicals derived from these cellular activities are known as biochemicals. Fuels produced these cellular activities are known as biofuels. Biofuels are a renewable alternative to petroleum based fuels. Biofuels can be substituted for any petroleum based fuel (e.g., gasoline, diesel, aviation fuel, heating oil, etc.). Biofuels can be derived from renewable sources, such as plant matter, animal matter, or even waste products. These renewable sources are collectively known as biomass. One advantage of biofuels over petroleum based fuels is that they do not require expensive and risky exploration or extraction. In addition, biofuels can be locally produced. Hence, they do not require transportation over long distances. Moreover, biofuels can be made directly without the need for expensive and energy intensive refining as is needed with refining crude petroleum. In other circumstances, the biofuel may require a limited and cost-effective level of refining. Furthermore, the use of biofuels improves the environment by reducing the amount of environmentally harmful emissions (e.g., greenhouse gases, air pollution, etc.) released during combustion. Finally, biofuels maintain a balanced carbon cycle because biofuels are produced from biomass, a renewable, natural resource. While the burning of biofuels will release carbon (e.g., as carbon dioxide), this carbon will be recycled during the production of biomass (e.g., the cultivation of crops) thereby balancing the carbon cycle unlike petroleum based fuels.

[0021] For similar reasons, biologically derived chemicals offer the same advantages as biofuels over petroleum based fuels. Biologically derived chemicals are a renewable alternative to petrochemicals. Biologically derived chemicals, such as hydrocarbons (e.g., alkanes, alkenes, or alkynes), fatty alcohols, esters, fatty acids, fatty aldehydes, and ketones are superior to petrochemicals because they are produced directly without extensive refining. Unlike petrochemicals, biologically derived chemicals do not need to be refined like crude petroleum to recover raw materials which must then be further processed to make more complex petrochemicals. Biologically derived chemicals are directly converted from biomass to the desired chemical product.

[0022] To reduce our reliance on petroleum based fuels, it would be desirable to produce olefins from other sources.

SUMMARY OF THE INVENTION

[0023] The invention is based, at least in part, on the discovery of a novel gene, orf880, which encodes an olefin-producing enzyme. Accordingly, in one aspect, the invention features an isolated polynucleotide comprising a nucleotide sequence having at least 70% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In some embodiments, the nucleotide sequence has at least about 75%, at least about
80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In some embodiments, the polynucleotide comprises the nucleotide sequence of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In other embodiments, the polynucleotide consists of the nucleotide sequence of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13.

0024] In another aspect, the invention features an isolated polynucleotide comprising a nucleotide sequence that hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof. In some embodiments, the nucleotide sequence hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof under low stringency, medium stringency, high stringency, or very high stringency conditions.

0025] In another aspect, the invention features an isolated polynucleotide comprising a sequence encoding a polypeptide comprising (i) the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 or (ii) the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide comprises the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more conservative amino acid substitutions. For example, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more amino acid substitutions, additions, insertions, or deletions.

0026] In some embodiments, the polypeptide has fatty acid decarboxylase activity. In another aspect, the invention features an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide having the same biological activity as a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14, wherein the polynucleotide comprises: (i) the nucleotide sequence of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or a fragment thereof or (ii) a nucleotide sequence that hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof under low stringency, medium stringency, high stringency, or very high stringency conditions.

0028] In some embodiments, the polynucleotide comprises a sequence encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14.

0029] In other embodiments, a polynucleotide described herein is isolated from a bacterium, plant, insect, yeast, fungus, or mammal. In some embodiments, the polynucleotide is isolated from a bacterium, for example, a Gram positive or Gram negative bacterium. In specific embodiments, the bacterium is a member of the genus *Jeotgalicoccus*, *Corynebacterium*, *Kokuria*, *Methylobacterium*, or *Bacillus*. For example, the bacterium is selected from the group consisting of *Jeotgalicoccus halotolerans*, *Jeotgalicoccus psychrophilus*, *Jeotgalicoccus sp. AICC 8456*, *Jeotgalicoccus pinnipedalis*, *Corynebacterium efficiens*, *Kokuria rhizophila*, *Methylobacterium populic*, and *Bacillus subtilis*.

0030] In other embodiments, the isolated polynucleotide includes an operably linked promoter.

0031] In another aspect, the invention features an isolated polypeptide that includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14, or a biologically active fragment thereof. In some embodiments, the polypeptide consists of the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 or a biologically active fragment thereof. In some embodiments, the fragment is at least about 25 amino acids in length, for example, at least about 30, at least about 40, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, or at least about 250 amino acids in length. In certain embodiments, the fragment has fatty acid decarboxylase activity.

0032] In yet another aspect, the invention features an isolated polypeptide that includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more conservative amino acid substitutions. For example, the polypeptide includes one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue.

0033] In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has fatty acid decarboxylase activity.

0034] In another aspect, the invention features an isolated polypeptide that includes the amino acid sequence F-X-E-[VI]-[RK]-R-X-Y-P-F-[F]-P-X-[LIV], where X is any amino acid and [F] stands for any amino acid except Phe; and wherein the polypeptide has fatty acid decarboxylase activity. In an alternate embodiment, the invention features an isolated polypeptide that includes the amino acid sequence P-X(6)-[I]-X(4)-[G]-[VI]-[HQM]-X-[MLV]-D-G-X(2)-H-X(2)-R-K, where X is any amino acid and [G] stands for any amino acid except Gly; wherein the polypeptide has fatty acid decarboxylase activity. In another embodiment, the invention features an isolated polypeptide that includes the amino acid sequence [AC]-[AG]-[HV]-[DE]-[HL]-X-N-X(2)-R-P-X-[VI]-X(3)-[FY]-X(2)-F-X.
(3)-A-[LMV]-X-[DE] where X is any amino acid; and wherein the polypeptide has fatty acid decarboxylase activity. In another embodiment, the invention features an isolated polypeptide that includes the amino acid sequence \( \text{[RK]}-\text{X}[6]-\text{[LMV]}-\text{X}[4]-\text{[AC]}-\text{[Ag]}-\text{[IV]}-\text{[DE]}-\text{[II]}\)-\( \text{X}[N] \)-\( \text{X}[2] \)-R-P-X-[VI]-A-X[3]-[FY]-X[2]-F-X[3]-A-[LMV]-X-[DE] \) where X is any amino acid and \([RK]\) stands for any amino acid except Arg and Lys; and wherein the polypeptide has fatty acid decarboxylase activity.

In another aspect, the invention features a recombinant vector that includes a polynucleotide described herein. In some embodiments, the vector is a plasmid. In other embodiments, the vector further includes a promoter operably linked to the polynucleotide, for example, a developmentally-regulated, an organelle-specific, a tissue-specific, an inducible, a constitutive, or a cell-specific promoter.

In some embodiments, the vector includes at least one sequence selected from the group consisting of: (a) a regulatory sequence operatively coupled to the polynucleotide; (b) a selection marker operatively coupled to the polynucleotide; (c) a marker sequence operatively coupled to the polynucleotide; (d) a purification moiety operatively coupled to the polynucleotide; (e) a secretion sequence operatively coupled to the polynucleotide; and (f) a targeting sequence operatively coupled to the polynucleotide.

In another aspect, the invention features a host cell comprising a recombinant vector described herein, for example, a recombinant vector that includes a polynucleotide described herein. In some embodiments, the host cell expresses a polypeptide encoded by the recombinant vector, for example, encoded by a polynucleotide sequence included in the recombinant vector. In some embodiments, the polynucleotide sequence is stably incorporated into the genomic DNA of the host cell. In particular embodiments, the expression of the polynucleotide sequence is under the control of a regulated promoter region.

In some embodiments, the host cell is a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, a filamentous fungi cell, or bacterial cell.

In some embodiments, the host cell is a Gram positive bacterial cell. In other embodiments, the host cell is a Gram negative bacterial host cell.

In some embodiments, the host cell is from a member of the genus Escherichia, Bacillus, Lacticobacillus, Rhodococcus, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophthora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Stenotrophomonas, Schizosaccharomyces, Yarrowia, or Streptomyces.

In particular embodiments, the host cell is a Bacillus lentes cell, a Bacillus brevis cell, a Bacillus stea rothermophilus cell, a Bacillus licheniformis cell, a Bacillus alkaliophilus cell, a Bacillus coagulans cell, a Bacillus circulans cell, a Bacillus pumilis cell, a Bacillus thuringiensis cell, a Bacillus clausii cell, a Bacillus megaterium cell, a Bacillus subtilis cell, or a Bacillus amyloliquefaciens cell.

In other embodiments, the host cell is a Trichoderma koningii cell, a Trichoderma viride cell, a Trichoderma reesei cell, a Trichoderma longibrachiatum cell, an Aspergillus awamori cell, an Aspergillus fumigatus cell, an Aspergillus foetidus cell, an Aspergillus nidulans cell, an Aspergillus niger cell, an Aspergillus oryzae cell, a Humicola insolens cell, a Humicola lanuginose cell, a Rhodococcus opacus cell, a Rhizomucor miehei cell, or a Mucor miehei cell.

In yet another embodiments, the host cell is a Streptomyces lividans cell or a Streptomyces murinus cell. In other embodiments, the host cell is an Actinomycetes cell.

In some embodiments, the host cell is a CHO cell, a COS cell, a VERO cell, a BHK cell, a HeLa cell, a Cv1 cell, an MDCK cell, a 293 cell, a 3T3 cell, or a PC12 cell.

In particular embodiments, the host cell is an E. coli cell, such as a strain B, a strain C, a strain K, or a strain W, E. coli cell.

In certain embodiments, the host cell produces an olefin, for example, a terminal olefin. In some embodiments, the olefin is secreted by the host cell.

In other embodiments, the host cell expresses a reduced level of an acyl-CoA synthase relative to a corresponding wild type host cell. In some embodiments, a gene encoding an acyl-CoA synthase is attenuated or deleted in the host cell. In certain embodiments, the gene is fasD, fasK, BH3103, yhlF, Pfl-4354, EAV15023, fasD1, fasD2, RFC_4074, fasDD35, fasDD22, fas1p, fas3p, IeCA, or the gene encoding the protein having the amino acid sequence Accession No. ZP_01644857. In particular embodiments, the gene is fasD.

In another aspect, the invention features a genetically engineered microorganism comprising an exogenous control sequence stably incorporated into the genomic DNA of the microorganism. In one embodiment, the control sequence is integrated upstream of a polynucleotide comprising a nucleotide sequence having at least about 70% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In some embodiments, the nucleotide sequence has at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In some embodiments, the nucleotide sequence is SEQ ID NO:1, 4, 6, 8, 9, 11, or 13.

In some embodiments, the polynucleotide is endogenous to the microorganism. In some embodiments, the microorganism produces an increased level of a polypeptide encoded by the polynucleotide relative to a wild-type microorganism. In other embodiments, the microorganism produces an increased level of an olefin relative to a wild-type microorganism. In some embodiments, the microorganism is a bacterium, such as a Leotagallicoccus bacterium.

In other embodiments, the microorganism expresses a reduced level of an acyl-CoA synthase relative to a corresponding wild type microorganism. In some embodiments, a gene encoding an acyl-CoA synthase is attenuated or deleted in the microorganism. In certain embodiments, the gene is fasD, fasK, BH3103, yhlF, Pfl-4354, EAV15023, fasD1, fasD2, RFC_4074, fasDD35, fasDD22, fas1p, fas3p, IeCA, or the gene encoding the protein having the amino acid sequence Accession No. ZP_01644857. In particular embodiments, the gene is fasD.

In another aspect, the invention features a polypeptide produced by a host cell described herein.

In yet another aspect, the invention features a method of producing an olefin. The method comprises culturing a host cell described herein, for example, a host cell comprising a nucleotide sequence described herein,
under conditions sufficient to allow expression of a polypeptide encoded by the nucleotide sequence.

[0053] In some embodiments, the nucleotide sequence has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In some embodiments, the nucleotide sequence is SEQ ID NO:1, 4, 6, 8, 9, 11, or 13.

[0054] In other embodiments, the nucleotide sequence hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof, for example, under low stringency, medium stringency, high stringency, or very high stringency conditions.

[0055] In other embodiments, the nucleotide sequence encodes a polypeptide comprising: (i) the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14; or (ii) the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide comprises the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more conservative amino acid substitutions. For example, the polypeptide comprises one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of a threonine with a serine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue. In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the polypeptide has fatty acid deoxyxylase activity.

[0056] In other embodiments, the nucleotide sequence encodes a polypeptide having the same biological activity as a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14. In some embodiments, the nucleotide sequence is SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or a fragment thereof. In other embodiments, the nucleotide sequence hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof; for example, under low stringency, medium stringency, high stringency, or very high stringency conditions. In some embodiments, the biological activity is fatty acid deoxyxylase activity.

[0057] In some embodiments, the nucleotide sequence encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14.

[0058] In some embodiments, the host cell is selected from the group consisting of a mammalian cell, plant cell, insect cell, yeast cell, fungus cell, filamentous fungi cell, and bacterial cell.

[0059] In some embodiments, the host cell is a Gram positive bacterial cell. In other embodiments, the host cell is a Gram negative bacterial cell.
acid substrate, a branched chain fatty acid substrate, or a fatty acid substrate that includes a cyclic moiety.

[0071] In other embodiments, the host cell expresses a reduced level of an acyl-CoA synthase relative to a corresponding wild type host cell. In some embodiments, a gene encoding an acyl-CoA synthase is attenuated or deleted in the host cell. In certain embodiments, the gene isfadD, faaK, BH3103, yhlL, PTI-4534, EAV15023, fadD1, fadD2, RPC-4074, faaDD35, fadDD22, faa1p, faa3p, lefA, or the gene encoding the protein having the amino acid sequence Accession No. ZP_01644857. In particular embodiments, the gene isfadD.

[0072] In some embodiments, the method further includes isolating the olefin from the host cell or from the culture medium. In other embodiments, the method further includes cracking or refining the olefin.

[0073] In another aspect, the invention features a method of making an olefin. The method includes contacting a biological substrate with an isolated polypeptide having the same biological activity as a polypeptide comprising the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14.

[0074] In some embodiments, the olefin is a terminal olefin. In some embodiments, the terminal olefin is a C3-C25 terminal olefin. For example, the terminal olefin is a C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or C25 terminal olefin. In some embodiments, the terminal olefin is nonadecene, methylnonadecene, heptadecene, methylheptadecene, or pentadecene.

[0075] In some embodiments, the terminal olefin is an unsaturated terminal olefin or a monounsaturated terminal olefin. In yet other embodiments, the terminal olefin is a straight chain terminal olefin, a branched chain terminal olefin, or a terminal olefin that includes a cyclic moiety.

[0076] In some embodiments, the biological substrate is a fatty acid substrate. In certain embodiments, the fatty acid substrate is a C4-C20 fatty acid substrate. For example, the fatty acid substrate is a C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or C25 fatty acid substrate. In particular embodiments, the fatty acid substrate is methyl eicosanoic acid, eicosanoic acid, methyl octadecanoic acid, stearic acid, or palmitic acid.

[0077] In some embodiments, the fatty acid substrate is an unsaturated fatty acid substrate, a monounsaturated fatty acid substrate, or a saturated fatty acid substrate. In other embodiments, the fatty acid substrate is a straight chain fatty acid substrate, a branched chain fatty acid substrate, or a fatty acid substrate that includes a cyclic moiety.

[0078] In another aspect, the invention features a method of making an olefin. The method includes contacting a biological substrate with an isolated polypeptide encoded by a nucleotide sequence described herein. In some embodiments, the nucleotide sequence has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO:1, 4, 6, 8, 9, 11, or 13. In such embodiments, the nucleotide sequence is SEQ ID NO:1, 4, 6, 8, 9, 11, or 13.

[0079] In other embodiments, the nucleotide sequence hybridizes to a complement of SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 or to a fragment thereof, for example, under low stringency, medium stringency, high stringency, or very high stringency conditions.

[0080] In some embodiments, the olefin is a terminal olefin. In some embodiments, the terminal olefin is a C3-C25 terminal olefin. For example, the terminal olefin is a C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or C25 terminal olefin. In some embodiments, the terminal olefin is nonadecene, methylheptadecene, heptadecene, methylheptadecene, or pentadecene.

[0081] In some embodiments, the terminal olefin is an unsaturated terminal olefin or a monounsaturated terminal olefin. In yet other embodiments, the terminal olefin is a straight chain terminal olefin, a branched chain terminal olefin, or a terminal olefin that includes a cyclic moiety.

[0082] In some embodiments, the biological substrate is a fatty acid substrate. In certain embodiments, the fatty acid substrate is a C6-C24 fatty acid substrate. For example, the fatty acid substrate is a C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or C25 fatty acid substrate. In particular embodiments, the fatty acid substrate is methyl eicosanoic acid, eicosanoic acid, methyl octadecanoic acid, stearic acid, or palmitic acid.

[0083] In some embodiments, the biological substrate is an unsaturated fatty acid substrate, a monounsaturated fatty acid substrate, or a saturated fatty acid substrate. In other embodiments, the fatty acid substrate is a straight chain fatty acid substrate, a branched chain fatty acid substrate, or a fatty acid substrate that includes a cyclic moiety.

[0084] In another aspect, the invention features a method of making an olefin. The method includes contacting a biological substrate with an isolated polypeptide described herein. In such embodiments, the polypeptide includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 or a biologically active fragment thereof. In some embodiments, the polypeptide consists of the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 or a biologically active fragment thereof. In some embodiments, the fragment is at least about 25 amino acids in length, for example, at least about 30, at least about 40, at least about 50, at least about 75, at least about 100, at least about 200, or at least about 250 amino acids in length. In certain embodiments, the fragment has fatty acid decarboxylase activity.

[0085] In other embodiments, the polypeptide includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more amino acid substitutions, additions, insertions, or deletions. In such embodiments, the polypeptide includes the amino acid sequence of SEQ ID NO:2, 5, 7, 10, 12, or 14 with one or more conservative amino acid substitutions. For example, the polypeptide includes one or more of the following conservative amino acid substitutions: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group, exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue.
In some embodiments, the polypeptide has about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more amino acid substitutions, additions, insertions, or deletions. In some embodiments, the isolated polypeptide has fatty acid decarboxylase activity.

[0086] In other embodiments, the polypeptide includes the amino acid sequence F-X-X-E-[VI]-[RK]-R-X-Y-P-F-[P]-P-X-[ILV]; and wherein the polypeptide has fatty acid decarboxylase activity. In an alternate embodiment, the polypeptide includes the amino acid sequence P-X(6)-[I]-X(4)-[G]-[VI]-[HQM]-X-[MLV]-D-G-X(2)-H-X(2)-R-K; and wherein the polypeptide has fatty acid decarboxylase activity. In another embodiment, the polypeptide includes the amino acid sequence [AC]-[AG]-[IV]-[DE]-[IL]-X-N-X(2)-R-P-X-[VI]-A-X(3)-[FY]-X(2)-F-X(3)-A-[LMV]-X-[DE]; and wherein the polypeptide has fatty acid decarboxylase activity. In another embodiment, the polypeptide includes the amino acid sequence [RK]-X(6)-[LMV]-X(4)-[AC]-[AG]-[IV]-[DE]-[IL]-X-N-X(2)-R-P-X-[VI]-A-X(3)-[FY]-X(2)-F-X(3)-A-[LMV]-X-[DE]; and wherein the polypeptide has fatty acid decarboxylase activity. In the sequences described herein, X is any amino acid, [AB] is A or B (e.g., [VI] is valine or isoleucine), [A] is any amino acid except A (e.g., [P] stands for any amino acid except Phe), X is any amino acid, and X(n) is X repeated n times (e.g., X(4) is X-X-X-X).

[0087] In some embodiments, the olefin is a terminal olefin. In some embodiments, the terminal olefin is a C8-C25 terminal olein. For example, the terminal olefin is a C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or C25 terminal olefin. In some embodiments, the terminal olefin is nonadecene, methylnonadecene, heptadecene, methylheptadecene, or pentadecene.

[0088] In some embodiments, the terminal olein is an unsaturated terminal olefin or a monounsaturated terminal olefin. In yet other embodiments, the terminal olein is a straight chain terminal olefin or a branched chain terminal olefin, or a terminal olefin that includes a cyclic.

[0089] In some embodiments, the biological substrate is a fatty acid substrate. In certain embodiments, the fatty acid substrate is a C8-C24 fatty acid substrate. For example, the fatty acid substrate is a C8, C9, C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, or a C24 fatty acid substrate. In particular embodiments, the fatty acid substrate is methyl eicosanoic acid, eicosanoic acid, methyl octadecanoic acid, stearic acid, or palmitic acid.

[0090] In some embodiments, the fatty acid substrate is an unsaturated fatty acid substrate, a monounsaturated fatty acid substrate, or a saturated fatty acid substrate. In other embodiments, the fatty acid substrate is a straight chain fatty acid substrate, a branched chain fatty acid substrate, or a fatty acid substrate that includes a cyclic moiety.

[0091] In another aspect, the invention features an olefin produced by any of the methods described herein. In particular embodiments, the olefin has a δ13C of about -15.4 or greater. For example, the olefin has a δ13C of about -15.4 to about -10.9, for example, about -13.92 to about -13.84. In other embodiments, the olefin has an f14C of at least about 1.003. For example, the olefin has an f14C of at least about 1.01 or at least about 1.5. In some embodiments, the olefin has an f14C of about 1.11 to about 1.124.

[0092] In another aspect, the invention features a biofuel that includes an olefin produced by any of the methods described herein. In particular embodiments, the olefin has a δ13C of about -15.4 or greater. For example, the olefin has a δ13C of about -15.4 to about -10.9, for example, about -13.92 to about -13.84. In other embodiments, the olefin has an f14C of at least about 1.003. For example, the olefin has an f14C of at least about 1.01 or at least about 1.5. In some embodiments, the olefin has an f14C of about 1.11 to about 1.124. In some embodiments, the biofuel is a biodiesel, gasoline, or jet fuel.

DEFINITIONS

[0093] Throughout the specification, a reference may be made using an abbreviated gene name or polypeptide name, but it is understood that such an abbreviated gene or polypeptide name represents the genus of genes or polypeptides. Such gene names include all genes encoding the same polypeptide and homologous polypeptides having the same physiological function. Polypeptide names include all polypeptides that have the same activity (e.g., that catalyze the same fundamental chemical reaction).

[0094] The accession numbers referenced herein are derived from the NCBI database (National Center for Biotechnology Information) maintained by the National Institutes of Health, U.S.A. The accession numbers are as provided in the database as of Aug. 28, 2008.

[0095] EC numbers are established by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The EC numbers referenced herein are derived from the KEGG Ligand database, maintained by the Kyoto Encyclopedia of Genes and Genomics, sponsored in part by the University of Tokyo. The EC numbers are as provided in the database as of March, 2008.

[0096] The articles “a” and “an” are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.

[0097] The term “about” is used herein to mean a value ±20% of a given numerical value. Thus, “about 60%” means a value of between 60±(20% of 60) (i.e., between 48 and 70).

[0098] As used herein, the term “attenuate” means to weaken, reduce or diminish. For example, a polypeptide can be attenuated by modifying the polypeptide to reduce its activity (e.g., by modifying a nucleotide sequence that encodes the polypeptide).

[0099] As used herein, the term “biodiesel” means a biofuel that can be a substitute of diesel, which is derived from petroleum. Biodiesel can be used in internal combustion diesel engines in either a pure form, which is referred to as “neat” biodiesel, or as a mixture in any concentration with petroleum-based diesel. Biodiesel can include esters or hydrocarbons, such as olefins (e.g., terminal olefins).

[0100] As used therein, the term “biofuel” refers to any fuel derived from biomass. Biofuels can be substituted for petroleum based fuels. For example, biofuels are inclusive of transportation fuels (e.g., gasoline, diesel, jet fuel, etc.), heating fuels, and electricity-generating fuels. Biofuels are a renewable energy source.

[0101] As used herein, the term “biomass” refers to a carbon source derived from biological material. Biomass can be converted into a biofuel. One exemplary source of biomass is plant matter. For example, corn, sugar cane, or switchgrass can be used as biomass. Another non-limiting
example of biomass is animal matter, for example cow manure. Biomass also includes waste products from industry, agriculture, forestry, and households. Examples of such waste products that can be used as biomass are fermentation waste, straw, lumber, sewage, garbage, and food leftovers. Biomass also includes sources of carbon, such as carbohydrates (e.g., monosaccharides, disaccharides, or polysaccharides).

[0102] As used herein, the phrase “carbon source” refers to a substrate or compound suitable to be used as a source of carbon for prokaryotic or simple eukaryotic cell growth. Carbon sources can be in various forms, including, but not limited to polymers, carbohydrates, acids, alcohols, aldehydes, ketones, amino acids, peptides, and gases (e.g., CO and CO2). These include, for example, various monosaccharides, such as glucose, fructose, mannose, and galactose; oligosaccharides, such as fructo-oligosaccharide and galacto-oligosaccharide; polysaccharides such as xylose and arabinose; disaccharides, such as sucrose, maltose, and turanose; cellulose material, such as methyl cellulose and sodium carboxymethyl cellulose; saturated or unsaturated fatty acid esters, such as succinate, lactate, and acetate; and alcohols, such as ethanol or mixtures thereof. The carbon source can also be a product of photosynthesis, including, but not limited to, glucose. A preferred carbon source is biomass. Another preferred carbon source is glucose.

[0103] As used herein, a “cloud point lowering additive” is an additive added to a composition to decrease or lower the cloud point of a solution.

[0104] As used herein, the phrase “cloud point of a fluid” means the temperature at which dissolved solids are no longer completely soluble. Below this temperature, solids begin precipitating as a second phase giving the fluid a cloudy appearance. In the petroleum industry, cloud point refers to the temperature below which a solidified material or other heavy hydrocarbon crystallizes in a crude oil, refined oil, or fuel to form a cloudy appearance. The presence of solidified materials influences the flowing behavior of the fluid, the tendency of the fluid to clog fuel filters, injectors, etc., the accumulation of solidified materials on cold surfaces (e.g., a pipeline or heat exchanger fouling), and the emulsion characteristics of the fluid with water.

[0105] A nucleotide sequence is “complementary” to another nucleotide sequence if each of the bases of the two sequences matches (i.e., is capable of forming Watson Crick base pairs). The term “complementary strand” is a strand that complementary with the term “complement”. The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand.

[0106] As used herein, the term “conditions sufficient to allow expression” means any conditions that allow a host cell to produce a desired product, such as a polypeptide or olefin described herein. Suitable conditions include, for example, fermentation conditions. Fermentation conditions can comprise many parameters, such as temperature ranges, levels of aeration, and media composition. Each of these conditions, individually and in combination, allows the host cell to grow. Exemplary culture media include broth or gels. Generally, the medium includes a carbon source, such as glucose, fructose, cellulose, or the like, that can be metabolized by a host cell directly. In addition, enzymes can be used in the medium to facilitate the mobilization (e.g., the depolymerization of starch or cellulose to fermentable sugars) and subsequent metabolism of the carbon source.

[0107] To determine if conditions are sufficient to allow expression, a host cell can be cultured, for example, for about 4, 8, 12, 24, 36, or 48 hours. During and/or after culturing, samples can be obtained and analyzed to determine if the conditions allow expression. For example, the host cells in the sample or the medium in which the host cells were grown can be tested for the presence of a desired product. When testing for the presence of a product, assays, such as, but not limited to, TLC, HPLC, GC/FID, GC/MS, LC/MS, MS, can be used.

[0108] It is understood that the polypeptides described herein may have additional conservative or non-essential amino acid substitutions, which do not have a substantial effect on the polypeptide functions. Whether or not a particular substitution will be tolerated (i.e., will not adversely affect desired biological properties, such as dehydrogenase activity) can be determined as described in Bowie et al., Science (1990) 247:1306-1310. A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

[0109] As used herein, “control element” means a transcriptional control element. Control elements include promoters and enhancers. The term “promoter element,” “promoter,” or “promoter sequence” refers to a DNA sequence that functions as a switch that activates the expression of a gene. If the gene is activated, it is said to be transcribed or participating in transcription. Transcription involves the synthesis of mRNA from the gene. A promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA. Control elements interact specifically with cellular proteins involved in transcription (Maniatis et al., Science 236:1237, 1987).

[0110] As used herein, the term “fatty acid” means a carboxylic acid having the formula RCOOH. R represents an aliphatic group, preferably an alkyl group. R can comprise about 4 and about 22 carbon atoms. Fatty acids can be saturated, monounsaturated, or polyunsaturated. In a preferred embodiment, the fatty acid is made from a fatty acid biosynthetic pathway.

[0111] As used herein, the term “fatty acid biosynthetic pathway” means a biosynthetic pathway that produces fatty acids. The fatty acid biosynthetic pathway includes fatty acid enzymes that can be engineered, as described herein, to produce fatty acids, and in some embodiments can be expressed with additional enzymes to produce fatty acids having desired carbon chain characteristics.

[0112] As used herein, “fatty acid enzyme” means any enzyme involved in fatty acid biosynthesis. Fatty acid enzymes can be expressed or overexpressed in host cells to produce fatty acids. Non-limiting examples of fatty acid enzymes include fatty acid synthases and thioesterases.
As used herein, “fraction of modern carbon” or “f_{tM},” has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOxl and HOxII, respectively. The fundamental definition relates to 0.95 times the $^{13}C/^{12}C$ isotope ratio HOxI (referenced to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (plant material), f_{tM} is approximately 1.1.

Calculations of “homology” between two sequences can be performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence that is aligned for comparison purposes is at least about 50%, preferably at least about 40%, more preferably at least about 50%, even more preferably at least about 60%, and even more preferably at least about 70%, at least about 80%, at least about 90%, or about 100% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid “identity” is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent homology between two amino acid sequences is determined using the Needleman and Wunsch (1970), J. Mol. Biol. 48:444-453, algorithm that has been incorporated into the GAP program in the GCG software package, using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent homology between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWFind program.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about which parameters should be applied to determine if a molecule is within a homology limitation of the claims) is a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

As used herein, a “host cell” is a cell used to produce a product described herein (e.g., an olefin described herein). A host cell can be modified to express or overexpress selected genes or to have attenuated expression of selected genes. Non-limiting examples of host cells include plant, animal, human, bacteria, yeast, or filamentous fungi cells.

As used herein, the term “hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions” describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either method can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6x sodium chloride/sodium citrate (SSC) at about 45° C., followed by two washes in 0.2xSSC, 0.1% SDS at least at 50° C. (the temperature of the washes can be increased to 55° C. for low stringency conditions); 2) medium stringency hybridization conditions in 6x SSC at about 45° C., followed by one or more washes in 0.2xSSC, 0.1% SDS at 60° C.; 3) high stringency hybridization conditions in 6x SSC at about 45° C., followed by one or more washes in 0.2xSSC, 0.1% SDS at 65° C.; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65° C., followed by one or more washes at 0.2xSSC, 1% SDS at 65° C. Very high stringency conditions (4) are the preferred conditions unless otherwise specified.

The term “isolated” as used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs or RNAs, respectively, that are present in the natural source of the nucleic acid. Moreover, by an “isolated nucleic acid” is meant to include nucleic acid fragments, which are not naturally occurring as fragments and would not be found in the natural state. The term “isolated” is also used herein to refer to polypeptides, which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. The term “isolated” as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques. The term “isolated” as used herein also refers to a nucleic acid or peptide that is substantially free of chemical precursors or other chemicals when chemically synthesized.

As used herein, the “level of expression of a gene in a cell” refers to the level of mRNA, pre-mRNA, mature transcript(s), transcript processing intermediates, mature mRNA(s), and degradation products encoded by the gene in the cell.

As used herein, the term “microorganism” means prokaryotic and eukaryotic microbial species from the domains Archaea, Bacteria and Eucarya, the latter including yeast and filamentous fungi, protozoa, algae, or higher Protista. The terms “microbial cells” (i.e., cells from microbes) and “microbes” are used interchangeably and refer to cells or small organisms that can only be seen with the aid of a microscope.

As used herein, the term “nucleic acid” refers to polynucleotides, such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides, ESTs, chromosomes, cDNAs, mRNAs, and rRNAs.

“Olefin” and “alkene” are used interchangeably herein, and “terminal olefin,” “cis-olefin,” and “terminal alkene” are used interchangeably herein.
As used herein, the term “openly linked” means that selected nucleotide sequence (e.g., encoding a polypeptide described herein) is in proximity with a promoter to allow the promoter to regulate expression of the selected DNA. In addition, the promoter is located upstream of the selected nucleotide sequence in terms of the direction of transcription and translation. By “openly linked” is meant that a nucleotide sequence and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence(s).

The term “or” is used herein to mean, and is used interchangeably with, the term “and/or,” unless context clearly indicates otherwise.

As used herein, “overexpress” means to express or cause to be expressed a nucleic acid, polypeptide, or hydrocarbon in a cell at a greater concentration than is normally expressed in a corresponding wild-type cell. For example, a polypeptide can be “overexpressed” in a recombinant host cell when the polypeptide is present in a greater concentration in the recombinant host cell compared to its concentration in a non-recombinant host cell of the same species.

As used herein, “partition coefficient” or “P,” is defined as the equilibrium concentration of a compound in an organic phase divided by the concentration at equilibrium in an aqueous phase (e.g., fermentation broth). In one embodiment of a bi-phasic system described herein, the organic phase is formed by the olefin during the production process. However, in some examples, an organic phase can be provided, such as by providing a layer of octane, to facilitate product separation. When describing a two phase system, the partition characteristics of a compound can be described as log P. For example, a compound with a log P of 1 would partition 10:1 to the organic phase. A compound with a log P of -1 would partition 1:10 to the organic phase. By choosing an appropriate fermentation broth and organic phase, an olefin with a high log P value can separate into the organic phase even at very low concentrations in the fermentation vessel.

As used herein, the term “purify,” “purified,” or “purification” means the removal or isolation of a molecule from its environment by, for example, isolation or separation. “Substantially purified” molecules are at least about 60% free, preferably at least about 75% free, and more preferably at least about 90% free from other components with which they are associated. As used herein, these terms also refer to the removal of contaminants from a sample. For example, the removal of contaminants can result in an increase in the percentage of olefins in a sample. For example, when olefins are produced in a host cell, the olefins can be purified by the removal of host cell proteins. After purification, the percentage of olefins in the sample is increased.

The terms “purify,” “purified,” and “purification” do not require absolute purity. They are relative terms. Thus, for example, when olefins are produced in host cells, a purified olefin is one that is substantially separated from other cellular components (e.g., nucleic acids, polypeptides, lipids, carbohydrates, or other hydrocarbons). In another example, a purified olefin preparation is one in which the olefin is substantially free from contaminants, such as those that might be present following fermentation. In some embodiments, an olefin is purified when at least about 50% by weight of a sample is composed of the olefin. In other embodiments, an olefin is purified when at least about 60%, 70%, 80%, 85%, 90%, 92%, 95%, 98%, or 99% or more by weight of a sample is composed of the olefin.

As used herein, the term “recombinant polypeptide” refers to a polypeptide that is produced by recombinant DNA techniques, wherein generally DNA encoding the expressed protein or RNA is inserted into a suitable expression vector and that is in turn used to transform a host cell to produce the polypeptide or RNA.

As used herein, the term “substantially identical” (or “substantially homologous”) is used to refer to a first amino acid or nucleotide sequence that contains a sufficient number of identical or equivalent (e.g., with a similar side chain, e.g., conserved amino acid substitutions) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have similar activities.

As used herein, the term “transfection” means the introduction of a nucleic acid (e.g., via an expression vector) into a recipient cell by nucleic acid-mediated gene transfer.

As used herein, “transformation” refers to a process in which a cell’s genotype is changed as a result of the cellular uptake of exogenous DNA or RNA. This may result in the transformed cell expressing a recombinant form of an RNA or polypeptide. In the case of antisense expression from the transferred gene, the expression of a naturally occurring form of the polypeptide is disrupted.

As used herein, a “transport protein” is a polypeptide that facilitates the movement of one or more compounds in and/or out of a cellular organelle and/or a cell.

As used herein, a “variant” of polypeptide X refers to a polypeptide having the amino acid sequence of peptide X in which one or more amino acid residues is altered. The variant may have conservative changes or nonconservative changes. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without affecting biological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The term “variant,” when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to that of a gene or the coding sequence thereof. This definition may also include, for example, “allelic,” “splice,” “species,” or “polymorphic” variants. A splice variant may have significant identity to a reference polynucleotide, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species.

As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of useful vector is an episome (i.e., a nucleic acid capable of extra-chromosomal replication). Useful vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as “expression vectors”. In general, expression vec-
tors of utility in recombinant DNA techniques are often in the form of “plasmids,” which refer generally to circular double stranded DNA loops that, in their vector form, are not bound to the chromosome. In the present specification, “plasmid” and “vector” are used interchangeably, as the plasmid is the most commonly used form of vector. However, also included are such other forms of expression vectors that serve equivalent functions and that become known in the art subsequently hereto.

[0137] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0138] Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0139] FIGS. 1A, 1B, 1C and 1D are GC/MS traces of olefins produced by Jeotgalicoccus sp. ATCC 8456 cells, Jeotgalicoccus halotolerans DSMZ 17274 cells, Jeotgalicoccus pinnipedalis DSMZ 17030 cells, and Jeotgalicoccus psychrophilus DSMZ 19085 cells, respectively.

[0140] FIGS. 2A and 2B are mass spectrometry fragmentation patterns of two α-olefins produced by Jeotgalicoccus ATCC 8456 cells. Compound A was identified as 1-nonadecene and compound B as 18-methyl-1-nonenadecene.

[0141] FIG. 3 is a schematic of a phylogenetic analysis of 16s rRNA of Jeotgalicoccus ATCC 8456.

[0142] FIGS. 4A and 4B are GC/MS traces of α-olefins produced by Jeotgalicoccus sp. ATCC 8456 cells upon feeding with eicosanoic acid (FIG. 4A) or stearic acid (FIG. 4B).

[0143] FIG. 5 is a GC/MS trace of α-olefins produced by cell free lysates of Jeotgalicoccus sp. ATCC 8456 cells.

[0144] FIG. 6 is a digital representation of an SDS-PAGE gel of final purified α-olefins producing protein fraction from Jeotgalicoccus sp. ATCC 8456 cells.

[0145] FIGS. 7A and 7B are the nucleotide- and corresponding amino acid sequences of Jeotgalicoccus sp. ATCC8456 orf880, respectively. FIG. 7C is the partial 16s rRNA sequence of Jeotgalicoccus sp. ATCC8456. FIGS. 7D and 7E are the codon-optimized nucleotide- and corresponding amino acid sequences of Jeotgalicoccus sp. ATCC8456 orf880, respectively. FIGS. 7F and 7G are the nucleotide- and corresponding amino acid sequences of Corynebacterium efficiens orf CE2459 (NP_739069), respectively. FIG. 7H is the nucleotide sequence of Kokuria rhizhoa orf KRH21570 (YP_001856010). FIGS. 7I and 7J are the codon-optimized nucleotide- and corresponding amino acid sequences of Kokuria rhizhoa orf KRH21570 (YP_001856010), respectively. FIGS. 7K and 7L are the codon-optimized nucleotide- and corresponding amino acid sequences of Methyllobacterium populi orf Mpop 1292 (YP_001923998). FIGS. 7M and 7N are the codon-optimized nucleotide- and corresponding amino acid sequences of Bacillus subtilis CYP 152A1 (NP_388902).

[0146] FIG. 8 is a GC/MS trace of α-olefins produced by E. coli upon expression of Jeotgalicoccus sp. 8456_orf880 and feeding of stearic acid.

[0147] FIG. 9 is a schematic of a bootstrap phylogenetic analysis of 8456_orf880 homologs using ClustalW.

[0148] FIGS. 10A-10Q represent a table identifying various genes that can be expressed, overexpressed or attenuated to increase production of particular fatty acid substrates.

[0149] FIG. 11 is a representation of a gel of PCR products from MG1655 wild type cells, ΔfadD::Cm cells, and ΔfadD cells.

[0150] FIG. 12A is a GC/MS trace of α-olefins produced by MG1655 ΔfadD cells transformed with empty vector. FIG. 12B is a GC/MS trace of α-olefins produced by MG1655 ΔfadD cells transformed with Jeotgalicoccus sp. 8456_orf880.

[0151] FIGS. 13A, 13B, and 13C are GC/MS traces of α-olefins produced by E. coli C41 ΔfadE cells transformed with empty vector (FIG. 13A), Corynebacterium efficiens YS-134 orf_CE2459 without stearic acid feeding (FIG. 13B), or with Corynebacterium efficiens YS-134 orf_CE2459 with stearic acid feeding (FIG. 13C).

[0152] FIGS. 14A, 14B, 14C, and 14D are GC/MS traces of α-olefins produced by E. coli C41 ΔfadE cells transformed with empty vector (FIG. 14A), empty vector with stearic acid feeding (FIG. 14B), Kokuria rhizhoa DC2201 orf KRH21570 without stearic acid feeding (FIG. 14C), or with Kokuria rhizhoa DC2201 orf KRH21570 with stearic acid feeding (FIG. 14D).

[0153] FIGS. 15A, 15B, and 15C are GC/MS traces of α-olefins produced by B. subtilis cells transformed with Jeotgalicoccus sp. ATCC8456_orf880 (FIG. 15A), Corynebacterium efficiens YS-134 orf_CE2459 (FIG. 15B), or empty vector (FIG. 15C).

[0154] FIG. 16 is a set of GC/MS traces of α-olefins produced in vitro by Corynebacterium efficiens ORF_CE2459.

[0155] FIGS. 17A, 17B, and 17C are GC/MS traces of α-olefins produced in vitro by purified Kokuria rhizhoa orf KRH21570 with hexadecanoic acid as substrate (FIG. 17A), by purified Kokuria rhizhoa orf KRH21570 with no substrate (FIG. 17B), and with no enzyme (FIG. 17C).

[0156] FIGS. 18A, 18B, and 18C are GC/MS traces of α-olefins produced in vitro by purified Methyllobacterium populi orf Mpop1292 with hexadecanoic acid as substrate (FIG. 18A), by purified Methyllobacterium populi orf Mpop1292 with no substrate (FIG. 18B), and with no enzyme (FIG. 18C).

[0157] FIG. 19 is a set of GC/MS traces of α-olefins produced by E. coli cells transformed with Bacillus subtilis P450_48 or empty vector.

[0158] FIG. 20 is a graph demonstrating the specific activity of 1-pentadecane and α(2)- and β(3)-hydroxy hexadecanoic acid production for ORF880, CYP152A1, and CYP152A1 Q85H using hexadecanoic acid as a substrate.

[0159] FIG. 21 is a graph demonstrating the ratio of decarboxylation to hydroxylation specific activities for ORF880, CYP152A1, and CYP152A1 Q85H.

[0160] FIG. 22 is a table identifying various P450 peroxigenases that can be expressed or overexpressed to produce α-olefins.
DETAILED DESCRIPTION

[0161] The invention provides compositions and methods of producing olefins from fatty acid substrates, in particular, terminal olefins. Such olefins are useful as a biofuel (e.g., substitutes for gasoline, diesel, jet fuel, etc.), specialty chemicals (e.g., lubricants, fuel additive, etc.), or feedstock for further chemical conversion (e.g., fuels, polymers, plastics, textiles, solvents, adhesives, etc.). The invention is based on the discovery of a novel gene, orf880, which encodes an olefin-producing enzyme. The polynucleotide sequence of orf880 and homologs of orf880 are presented as SEQ ID NO:1, 4, 6, 8, 9, 11, or 13 in FIG. 7, and the corresponding amino acid sequences are presented as SEQ ID NO:2, 5, 7, 10, 12, or 14 in FIG. 7. Using the methods described herein, olefins can be prepared using orf880, orf880 homologs, or variants thereof utilizing host cells or cell-free methods.

[0162] ORF880 Polynucleotide Variants

[0163] The methods and compositions described herein include the orf880 polynucleotide sequence (SEQ ID NO:1) as well as orf880 homologs depicted in FIG. 7 (SEQ ID NO: 4, 6, 8, 9, 11, or 13) and polynucleotide variants thereof. The variants can be naturally occurring or created in vitro. In particular, such variants can be created using genetic engineering techniques, such as site directed mutagenesis, random chemical mutagenesis, Exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives can be created using chemical synthesis or modification procedures.

[0164] Methods of making variants are well known in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids that encode polypeptides having characteristics that enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the sequence obtained from the natural isolate are generated and characterized. Typically, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

[0165] For example, variants can be created using error prone PCR (see, e.g., Leung et al., _Technique_ 1:11-15, 1989; and Caldwell et al., _PCR Methods Appl._ 2:28-33, 1992). In error prone PCR, PCR is performed under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. Briefly, in such procedures, nucleic acids to be mutated (e.g., an orf880 polynucleotide sequence), are mixed with PCR primers, reaction buffer, MgCl₂, MnCl₂, Taq polymerase, and an appropriate concentration of dNTPs for achieving a high rate of point mutation along the entire length of the PCR product. For example, the reaction can be performed using 20 nmol of nucleic acid to be mutated (e.g., an orf880 polynucleotide sequence), 50 pmol of each PCR primer, a reaction buffer comprising 50 mM KCl, 10 mM Tris HCl (pH 8.3), and 0.01% gelatin, 7 mM MgCl₂, 0.5 mM MnCl₂, 5 units of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM dTTP. PCR can be performed for 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min. However, it will be appreciated that these parameters can be varied as appropriate. The mutated nucleic acids are then cloned into an appropriate vector and the activities of the polypeptides encoded by the mutated nucleic acids are evaluated.

[0166] Variants can also be created using oligonucleotide directed mutagenesis to generate site-specific mutations in any cloned DNA of interest. Oligonucleotide mutagenesis is described in, for example, Reithauser-Olson et al., _Science_ 241:53-57, 1988. Briefly, in such procedures a plurality of double stranded oligonucleotides bearing one or more mutations to be introduced into the cloned DNA are synthesized and inserted into the cloned DNA to be mutagenized (e.g., an orf880 polynucleotide sequence). Clones containing the mutagenized DNA are recovered, and the activities of the polypeptides they encode are assessed.

[0167] Another method for generating variants is assembly PCR. Assembly PCR involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction. Assembly PCR is described in, for example, U.S. Pat. No. 5,965,408.

[0168] Still another method of generating variants is sexual PCR mutagenesis. In sexual PCR mutagenesis, forced homologous recombination occurs between DNA molecules of different, but highly related, DNA sequence in vitro as a result of random fragmentation of the DNA molecule based on sequence homology. This is followed by the crossover by primer extension in a PCR reaction. Sexual PCR mutagenesis is described in, for example, Stemmer, _PNAS_ USA 91:10747-10751, 1994.

[0169] Variants can also be created by in vivo mutagenesis. In some embodiments, random mutations in a nucleic acid sequence are generated by propagating the sequence in a bacterial strain, such as an _E. coli_ strain, which carries mutations in one or more of the DNA repair pathways. Such “mutator” strains have a higher random mutation rate than that of a wild-type strain. Propagating a DNA sequence (e.g., an orf880 polynucleotide sequence) in one of these strains will eventually generate random mutations within the DNA. Mutator strains suitable for use for in vivo mutagenesis are described in, for example, _PCT Publication No._ WO 91/16427.

[0170] Variants can also be generated using cassette mutagenesis. In cassette mutagenesis, a small region of a double stranded DNA molecule is replaced with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains a completely and/or partially randomized native sequence.

[0171] Recursive ensemble mutagenesis can also be used to generate variants. Recursive ensemble mutagenesis is an algorithm for protein engineering (i.e., protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. Recursive ensemble mutagenesis is described in, for example, Arkin et al., _PNAS_ USA 89:7811-7815, 1992.

[0172] In some embodiments, variants are created using exponential ensemble mutagenesis. Exponential ensemble mutagenesis is a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Exponential ensemble mutagenesis is
described in, for example, Delegrave et al., Biotech. Res. 11:1548-1552, 1993. Random and site-directed mutagenesis are described in, for example, Arnold, Curr. Opin. Biotech. 4:450-455, 1993.

[0173] In some embodiments, variants are created using shuffling procedures wherein portions of a plurality of nucleic acids that encode distinct polypeptides are fused together to create chimeric nucleic acid sequences that encode chimeric polypeptides as described in, for example, U.S. Pat. Nos. 5,965,408 and 5,939,250.

[0174] Polynucleotide variants also include nucleic acid analogs. Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxuryridine for deoxymethylidine and 5-methyl-2-deoxycytidilne or 5-bromo-2-deoxycytidine for deoxycytidin. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six-membered, morpholino ring, or peptide nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. (See, e.g., Summerton et al., Antisense Nucleic Acid Drug Dev. (1997) 7:187-195; and Hyrup et al., Bioorg. Med. Chem. (1996) 4:5-23.) In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorothioate backbone, a phosphoroumestrier backbone, or an alkyl phosphorostriester backbone.

ORF880 Polypeptide Variants

[0175] The methods and compositions described herein also include the ORF880 amino acid sequence (SEQ ID NO:2) depicted in FIG. 7, homologs of ORF880 (SEQ ID NO:5, 7, 10, 12, or 14) depicted in FIG. 7, and variants thereof. ORF880 polypeptide variants can be variants in which one or more amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue). Such substituted amino acid residue may or may not be one encoded by the genetic code.

[0176] Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of similar characteristics. Typical conservative substitutions are the following replacements: replacement of an aliphatic amino acid, such as alanine, valine, leucine, and isoleucine, with another aliphatic amino acid; replacement of a serine with a threonine or vice versa; replacement of an acidic residue, such as aspartic acid and glutamic acid, with another acidic residue; replacement of a residue bearing an amide group, such as asparagine and glutamine, with another residue bearing an amide group; exchange of a basic residue, such as lysine and arginine, with another basic residue; and replacement of an aromatic residue, such as phenylalanine and tyrosine, with another aromatic residue.

[0177] Other polypeptide variants are those in which one or more amino acid residues include a substituent group. Still other polypeptide variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (e.g., polyethylene glycol).

[0178] Additional polypeptide variants are those in which additional amino acids are fused to the polypeptide, such as a leader sequence, a secretary sequence, a proprotein sequence, or a sequence which facilitates purification, enrichment, or stabilization of the polypeptide.

[0179] In some instances, the polypeptide variants retain the same biological function as ORF880 (e.g., retain decarboxylase activity) and have amino acid sequences substantially identical thereto.

[0180] In other instances, the polypeptide variants have at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or more than about 95% homology to the amino acid sequence depicted in FIG. 7 (SEQ ID NO:2, 5, 7, 10, 12, or 14). In another embodiment, the polypeptide variants include a fragment comprising at least about 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof.

[0181] The polypeptide variants or fragments thereof can be obtained by isolating nucleic acids encoding them using techniques described herein or by expressing synthetic nucleic acids encoding them. Alternatively, polypeptide variants or fragments thereof can be obtained through biochemical enrichment or purification procedures. The sequence of polypeptide variants or fragments can be determined by proteolytic digestion, gel electrophoresis, and/or microsequencing. The sequence of the polypeptide variants or fragments can then be compared to the amino acid sequence depicted in FIG. 7 (SEQ ID NO:2, 5, 7, 10, 12, or 14) using any of the programs described herein.

[0182] The polypeptide variants and fragments thereof can be assayed for olefin-producing activity using routine methods. For example, the polypeptide variants or fragment can be contacted with a substrate (e.g., a fatty acid substrate) under conditions that allow the polypeptide variant to function. A decrease in the level of the substrate or an increase in the level of an olefin can be measured to determine olefin-producing activity.

ORF880 Motifs

[0183] By using bioinformatics, amino acid motifs can be designed by identifying conserved regions of the fatty acid decarboxylase (e.g., Geotrichum candidum sp. 8456 orf880). These amino acid motifs can be designed by methods well known in the art, such as bioinformatics, phylogenetic study, and/or protein alignments followed by visual inspection of the protein sequences. These amino acid motifs can then be used to identify proteins that have similar biological functions as the fatty acid decarboxylase. Several programs well known in the art can use the amino acid motifs to identify proteins that belong to the family of functional proteins.

[0184] Based on the experimental data, it was predicted that organisms that contain polypeptide sequences containing these amino acid motifs may be functional fatty acid decarboxylases.
In the sequences described herein, [AB] is A or B (e.g., [VI] is valine or isoleucine), {A} is any amino acid except A (e.g., [F] stands for any amino acid except Phe), X represents any amino acid (e.g., any naturally occurring amino acid), and X(n) is X repeated n times (e.g., X(4) is X-X-X-X).

0185 The invention is directed to an isolated nucleic acid encoding a polypeptide comprising a fatty acid decarboxylase amino acid motif sequence. For example, the isolated nucleic acid can encode a polypeptide comprising an amino acid sequence selected from the motifs described herein.

0186 Preferably, the isolated nucleic acid encodes a polypeptide of no more than about 700, 650, 600, 550, 500, or 450 amino acid residues comprising a fatty acid decarboxylase amino acid motif sequence.

Anti-Orf880 Antibodies

0187 The Orf880 polypeptides described herein can be used to produce antibodies directed against Orf880 polypeptides. Such antibodies can be used, for example, to detect the expression of an Orf880 polypeptide using methods known in the art. The antibody can be, e.g., a polyclonal antibody; a monoclonal antibody or antigen binding fragment thereof; a modified antibody such as a chimeric antibody, reshaped antibody, humanized antibody, or fragment thereof (e.g., Fab', Fab, Fab(ab')2); or a biosynthetic antibody, e.g., a single chain antibody, single domain antibody (DAbs), Fv, single chain Fv (scFv), or the like.

0188 Methods of making and using polyclonal and monoclonal antibodies are described, e.g., in Harlow et al., *Using Antibodies: A Laboratory Manual: Portable Protocol I.* Cold Spring Harbor Laboratory (Dec. 1, 1998). Methods for making modified antibodies and antibody fragments (e.g., chimeric antibodies, reshaped antibodies, humanized antibodies, or fragments thereof, e.g., Fab', Fab, Fab(ab')2 fragments); or biosynthetic antibodies (e.g., single chain antibodies, single domain antibodies (DAbs), Fv, single chain Fv (scFv), and the like), are known in the art and can be found, e.g., in Zola, *Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives*, Springer Verlag (Dec. 15, 2000; 1st edition).

Fatty Acid Substrates

0189 The compositions and methods described herein can be used to produce olefins (e.g., terminal olefins) from fatty acid substrates. While not wishing to be bound by theory, it is believed that the polypeptides described herein produce olefins from fatty acid substrates via a decarboxylation mechanism. Thus, olefins having particular branching patterns, levels of saturation, and carbon chain length can be produced from fatty acid substrates having those particular characteristics. Accordingly, each step within a fatty acid biosynthetic pathway can be modified to produce or overproduce a fatty acid substrate of interest. For example, known genes involved in the fatty acid biosynthetic pathway can be expressed, overexpressed, or attenuated in host cells to produce a desired fatty acid substrate (see, e.g., PCT/US08/057888). Example genes are provided in FIG. 10. It is recognized, however, that the carbon chain of an olefin produced using the methods described herein will have one less carbon than the fatty acid substrate from which it was produced.

Synthesis of Fatty Acid Substrates

0190 Fatty acid synthase (FAS) is a group of polypeptides that catalyze the initiation and elongation of acyl chains (Marrakchi et al., *Biochemical Society, 30*:1050-1055, 2002). The acyl carrier protein (ACP) along with the enzymes in the FAS pathway control the length, degree of saturation, and branching of the fatty acids produced. The fatty acid biosynthetic pathway involves the precursors acetyl-CoA and malonyl-CoA. The steps in this pathway are catalyzed by enzymes of the fatty acid biosynthesis (fab) and acetyl-CoA carboxylase (acc) gene families (see, e.g., Heath et al., *Prog. Lipid Res. 40*(6):467-97 (2001)).

0191 Host cells can be engineered to express fatty acid substrates by recombinantly expressing or overexpressing acetyl-CoA and/or malonyl-CoA synthase genes. For example, to increase acetyl-CoA production, one or more of the following genes can be expressed in a host cell: pdh, panK, aceEF, encoding the E1p dehydrogenase component and the E2p dihydrolipoamide acyltransferase component of the pyruvate and 2-oxoglutarate dehydrogenase complexes, fabH, fabD, fabG, acpP, and fabE. Example GenBank accession numbers for these genes are: pdh (BAB34380, AAC73227, AAC73226), panK (also known as coaA, AAC76952), aceEF (AAC73227, AAC73226), fabH (AAC74175), fabD (AAC74176), fabG (AAC74177), acpP (AAC74178), fabF (AAC74179). Additionally, the expression levels of fabE, gspa, ldhA, pdb, adhE, pta, poxB, ackA, and/or ackB can be attenuated or knocked-out in an engineered host cell by transformation with conditionally replicative or non-replicative plasmids containing null or deletion mutations of the corresponding genes or by substituting promoter or enhancer sequences. Example GenBank accession numbers for these genes are: fabE (AAC73325), gspa (AAC76632), ldhA (AAC74462), pdbB (AAC73989), adhE (AAC74323), pta (AAC73557), poxB (AAC7358), ackA (AAC73556), and ackB (BAB91430). The resulting host cells will have increased acetyl-CoA production levels when grown in an appropriate environment.

0192 Malonyl-CoA overexpression can be effected by introducing accABC (e.g., accession number AAC73296, EC 6.4.1.2) into a host cell. Fatty acids can be further overexpressed in host cells by introducing into the host cell a DNA sequence encoding a lipase (e.g., accession numbers CAA89087, CAA98876).

0193 In addition, inhibiting PslB can lead to an increase in the levels of long chain acyl-ACP, which will inhibit early steps in the pathway (e.g., accABC, fabH, and fabI). The PslB (e.g., accession number AAC77001) D311E mutation can be used to increase the amount of available acyl-CoA.

0194 In addition, a host cell can be engineered to overexpress a fabA gene (suppressor of fabA, e.g., accession number AAN79592) to increase production of monounsaturated fatty acids (Rock et al., *J. Bacteriology* 178:5382-5387, 1996).

0195 In some instances, host cells can be engineered to express, overexpress, or attenuate expression of a thio-
esterase to increase fatty acid substrate production. The chain length of a fatty acid substrate is controlled by thiosterase. In some instances, a text or a 1 gene can be overexpressed. In other instances, C12 fatty acids can be produced by attenuating thiosterase C12 (e.g., accession numbers AAC73596 and P00750), which uses C12 ACP, and expressing thiosterase C10 (e.g., accession number Q59133), which uses C10 ACP. This results in a relatively homogeneous population of fatty acids that have a carbon chain length of 10. In yet other instances, C12 fatty acids can be produced by attenuating endogenous thiosterases that produce non-C12 fatty acids and expressing the C10-ACP, that use C12 ACP (for example, accession number Q59473). In some situations, C12 fatty acids can be produced by expressing thiosterases that use C12 ACP (for example, accession number Q41635) and attenuating thiosterases that produce non-C12 fatty acids. Acetyl-CoA, malonyl-CoA, and fatty acid overproduction can be verified using methods known in the art, for example, by using radioactive precursors, HPLC, and GC-MS subsequent to cell lysis. Non-limiting examples of thiosterases that can be used in the methods described herein are listed in Table 1.

### TABLE 1

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Source Organism</th>
<th>Gene</th>
<th>Preferential product produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAC73596</td>
<td>E. coli</td>
<td>testA without leader sequence</td>
<td>C14-16</td>
</tr>
<tr>
<td>AAC73555</td>
<td>E. coli</td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>Q41635, AAA4215</td>
<td>Agrobacterium</td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>Q59133; AAC9206G</td>
<td>Cuphea hookerana</td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>AAC49269; AAC72881</td>
<td></td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>Q59473; AAC49151</td>
<td>Cuphea hookerana</td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>CAA85384</td>
<td>Arachis hypogaea</td>
<td>testB</td>
<td>C12-14</td>
</tr>
<tr>
<td>NP_199417; NP_193041</td>
<td></td>
<td>testA</td>
<td>C12-14</td>
</tr>
<tr>
<td>CAC3916</td>
<td>Bradyrhizobium japonicum</td>
<td>testA</td>
<td>C12-14</td>
</tr>
<tr>
<td>AAC72883</td>
<td>Cuphea hookerana</td>
<td>testA</td>
<td>C12-14</td>
</tr>
<tr>
<td>AAL75611</td>
<td>Helianthus annuus</td>
<td>testA</td>
<td>C12-14</td>
</tr>
</tbody>
</table>

*Mayer et al., BMC Plant Biology 7: 1-11, 2007*

**Formation of Branched Olefins**

[0196] Olefins can be produced that contain branch points by using branched fatty acids as substrates. For example, although *E. coli* naturally produces straight chain fatty acids (sFAs), *E. coli* can be engineered to produce branched chain fatty acids (bFAs) by introducing and expressing or overexpressing genes that provide branched precursors in the *E. coli* (e.g., bkd, ilv, icm, and fab gene families). Additionally, a host cell can be engineered to express or overexpress genes encoding proteins for the elongation of bFAs (e.g., ACP, FabF, etc.) and/or to delete or attenuate the corresponding host cell genes that normally lead to sFAs.

[0197] The first step in forming bFAs is the production of the corresponding α-keto acids by a branched-chain amino acid amidotransferase. Host cells may endogenously include genes encoding such enzymes or such genes can be recombinantly introduced. *E. coli*, for example, endogenously expresses such an enzyme, IlvE (EC 2.6.1.42; GenBank accession YP_026247). In some host cells, a heterologous branched-chain amino acid amidotransferase may not be expressed. However, *E. coli* ENB or any other branched-chain amino acid amidotransferase (e.g., IlvE from *Lactococcus lactis* (GenBank accession AAF34406), IlvE from *Pseudomonas putida* (GenBank accession NP_745648), or IlvE from *Streptomyces coelicolor* (GenBank accession NP_629657)), if not endogenous, can be introduced.

[0198] The second step is the oxidative decarboxylation of the α-keto acids to the corresponding branched-chain acyl-CoA. This reaction can be catalyzed by a branched-chain α-keto acid dehydrogenase complex (bkd; EC 1.2.4.4,) (Denova et al., *J. Bacteriol.* 177:3504, 1995), which consists of E1α/β (decarboxylase), E2 (dihydrolipoyl transacetylase), and E3 (dihydrolipoyl dehydrogenase) subunits. These branched-chain α-keto acid dehydrogenase complexes are similar to pyruvate and α-ketoglutarate dehydrogenase complexes. Any microorganism that possesses bFAs and/or grows on branched-chain amino acids can be used as a source to isolate bkd genes for expression in host cells, for example, *E. coli*. Furthermore, *E. coli* has the E3 component as part of its pyruvate dehydrogenase complex (Ipd, EC 1.8.1.4, GenBank accession NP_414658). Thus, it can be sufficient to express only the E1α/β and E2 bkd genes. Table 2 lists non-limiting examples of bkd genes from several microorganisms that can be recombinantly introduced and expressed in a host cell to provide branched-chain acyl-CoA precursors.

### TABLE 2

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GenBank Accession #</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>bkdA1 (E1α)</td>
<td>NP_628006</td>
</tr>
<tr>
<td></td>
<td>bkdB1 (E1β)</td>
<td>NP_628005</td>
</tr>
<tr>
<td><em>Streptomyces avermitilis</em></td>
<td>bkdA1 (E1α)</td>
<td>BAC72074</td>
</tr>
<tr>
<td></td>
<td>bkdB1 (E1β)</td>
<td>BAC72075</td>
</tr>
<tr>
<td></td>
<td>bkdF1 (E1α)</td>
<td>BAC72085</td>
</tr>
<tr>
<td></td>
<td>bkdF1 (E1β)</td>
<td>BAC72087</td>
</tr>
<tr>
<td><strong>Bacillus subtilis</strong></td>
<td>bkdA1 (E1α)</td>
<td>NP_390288</td>
</tr>
<tr>
<td></td>
<td>bkdA1 (E1β)</td>
<td>NP_390286</td>
</tr>
<tr>
<td></td>
<td>bkdB1 (E1β)</td>
<td>NP_390281</td>
</tr>
<tr>
<td><strong>Pseudomonas putida</strong></td>
<td>bkdA1 (E1α)</td>
<td>AAA5614</td>
</tr>
<tr>
<td></td>
<td>bkdA2 (E1β)</td>
<td>AAA5615</td>
</tr>
<tr>
<td></td>
<td>bkdC1 (E2)</td>
<td>AAA5617</td>
</tr>
</tbody>
</table>
In another example, isobutyryl-CoA can be made in a host cell, for example in *E. coli*, through the coexpression of a crotonyl-CoA reductase (Ccr, EC 1.6.5.5, 1.1.1.1) and isobutyryl-CoA mutase (large subunit IcmA, EC 5.4.99.2; small subunit IcmB, EC 5.4.99.2) (Han and Reynolds, *J. Bacteriol.* 179:5157, 1997). Crotonyl-CoA is an intermediate in fatty acid biosynthesis in *E. coli* and other microorganisms. Non-limiting examples of ccr and icm genes from selected microorganisms are listed in Table 3.

TABLE 3

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GenBank Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>Ccr</td>
<td>NP_630556</td>
</tr>
<tr>
<td></td>
<td>icmA</td>
<td>NP_629554</td>
</tr>
<tr>
<td></td>
<td>icmB</td>
<td>NP_630904</td>
</tr>
<tr>
<td><em>Streptomyces cinnamoneus</em></td>
<td>ccr</td>
<td>AAC5915</td>
</tr>
<tr>
<td></td>
<td>icmA</td>
<td>AAC08713</td>
</tr>
<tr>
<td></td>
<td>icmB</td>
<td>AI246005</td>
</tr>
</tbody>
</table>

In addition to expression of the bkd genes, the initiation of brFA biosynthesis utilizes β-ketoacyl-acyl-carrier-protein synthase III (FabH, EC 2.3.1.41) with specificity for branched chain acyl-CoAs (Li et al., *J. Bacteriol.* 187:3795-3799, 2005). Non-limiting examples of such FabH enzymes are listed in Table 4. fabH genes that are involved in fatty acid biosynthesis of any brFA-containing microorganism can be expressed in a host cell. The Bkd and FabH enzymes from host cells that do not naturally make brFA may not support brFA production. Therefore, bkd and fabH can be expressed recombinantly. Vectors containing the bkd and fabH genes can be inserted into such a host cell. Similarly, the endogenous level of Bkd and FabH production may not be sufficient to produce brFA. In this case, they can be overexpressed. Additionally, other components of the fatty acid biosynthesis pathway can be expressed or overexpressed, such as acyl carrier proteins (ACPs) and β-ketoacyl-acyl-carrier-protein synthase II (fabF, EC 2.3.1.41) (non-limiting examples of candidates are listed in Table 4). In addition to expressing these genes, some genes in the endogenous fatty acid biosynthesis pathway can be attenuated in the host cell (e.g., the *E. coli* genes fabH (GenBank accession # NP_415609) and/or fabF (GenBank accession # NP_415613)).

TABLE 4

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GenBank Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>fabH</td>
<td>NP_626634</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>NP_626635</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_626636</td>
</tr>
<tr>
<td><em>Streptomyces avermitilis</em></td>
<td>fabH</td>
<td>NP_823460</td>
</tr>
<tr>
<td></td>
<td>fabC3 (ACP)</td>
<td>NP_823467</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_823468</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>fabH</td>
<td>NP_389015</td>
</tr>
<tr>
<td></td>
<td>fabB</td>
<td>NP_388908</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>NP_389474</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_389016</td>
</tr>
<tr>
<td><em>Stenotrophomonas maltophilia</em></td>
<td>SmalDRAFT_0818</td>
<td>ZP_01640359</td>
</tr>
</tbody>
</table>

Formation of Cyclic Olefins

Cyclic olefins can be produced by using cyclic fatty acids as substrates. To produce cyclic fatty acid substrates, genes that provide cyclic precursors (e.g., the ans, chr, and plm gene families) can be introduced into the host cell and expressed to allow initiation of fatty acid biosynthesis from cyclic precursors. For example, to convert a host cell, such as *E. coli*, into one capable of synthesizing w-cyclic fatty acids (cyFA), a gene that provides the cyclic precursor cyclohexylcarbonyl-CoA (CHC-CoA) (Cropp et al., *Nature Biotech.* 18:980-983, 2000) can be introduced and expressed in the host cell. Non-limiting examples of genes that provide CHC-CoA in *E. coli* include: ansJ, ansK, ansL, chmA, and ansM from the ansatrienin gene cluster of *Streptomyces collium* (Chen et al., *Eur. J. Biochem.* 261:98-107, 1999) or plmJ, plmK, plmL, chmA, and plmM from the phthioceramycin B gene cluster of *Streptomyces* sp. HK803 (Palanippan et al., *J. Biol. Chem.* 278:35552-35557, 2003) together with the chmA gene (Patton et al., *Biochem.* 39:7595-7604, 2000) from *S. collins*, *S. avermitilis*, or *S. coelicolor* (see Table 5). The genes listed in Table 4 can then be expressed to allow initiation and elongation of w-cyclic fatty acids. Alternatively, the homologous genes can be isolated from microorganisms that make cyFA and expressed in a host cell (e.g., *E. coli*).

TABLE 5

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GenBank Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces collium</em></td>
<td>ansJ</td>
<td>U72144*</td>
</tr>
<tr>
<td></td>
<td>ansL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>chmA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ansM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>chicB</td>
<td></td>
</tr>
<tr>
<td><em>Streptomyces</em> sp. HK803</td>
<td>plmJ</td>
<td>AAQ84158</td>
</tr>
<tr>
<td></td>
<td>plmK</td>
<td>AAQ84159</td>
</tr>
<tr>
<td></td>
<td>plmL</td>
<td>AAQ84160</td>
</tr>
<tr>
<td></td>
<td>chmA</td>
<td>AAQ84161</td>
</tr>
<tr>
<td></td>
<td>plmM</td>
<td>AAQ84162</td>
</tr>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>chicB</td>
<td>NP_629292</td>
</tr>
<tr>
<td></td>
<td>chicL</td>
<td>NP_629292</td>
</tr>
<tr>
<td><em>Streptomyces avermitilis</em></td>
<td>chicB</td>
<td>NP_629292</td>
</tr>
</tbody>
</table>

*Only chicB is annotated in GenBank entry U72144, and ILK are according to Chen et al. (Eur. J. Biochem. 261:98-107, 1999).

In another example, isobutyryl-CoA can be made in a host cell, for example in *E. coli*, through the coexpression of a crotonyl-CoA reductase (Ccr, EC 1.6.5.5, 1.1.1.1) and isobutyryl-CoA mutase (large subunit IcmA, EC 5.4.99.2; small subunit IcmB, EC 5.4.99.2) (Han and Reynolds, *J. Bacteriol.* 179:5157, 1997). Crotonyl-CoA is an intermediate in fatty acid biosynthesis in *E. coli* and other microorganisms. Non-limiting examples of ccr and icm genes from selected microorganisms are listed in Table 3.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GenBank Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>fabH</td>
<td>NP_626634</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>NP_626635</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_626636</td>
</tr>
<tr>
<td><em>Streptomyces avermitilis</em></td>
<td>fabH</td>
<td>NP_823460</td>
</tr>
<tr>
<td></td>
<td>fabC3 (ACP)</td>
<td>NP_823467</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_823468</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>fabH</td>
<td>NP_389015</td>
</tr>
<tr>
<td></td>
<td>fabB</td>
<td>NP_388908</td>
</tr>
<tr>
<td></td>
<td>ACP</td>
<td>NP_389474</td>
</tr>
<tr>
<td></td>
<td>fabF</td>
<td>NP_389016</td>
</tr>
<tr>
<td><em>Stenotrophomonas maltophilia</em></td>
<td>SmalDRAFT_0818</td>
<td>ZP_01640359</td>
</tr>
</tbody>
</table>

The genes listed in Table 4 (fabH, ACP, and fabF) allow initiation and elongation of w-cyclic fatty acids because they have broad substrate specificity. If the coexpression of any of these genes with the genes listed in Table 5 does not yield cyFA, then fabH, ACP, and/or fabF homologs from microorganisms that make cyFA (e.g.,
those listed in Table 6) can be isolated (e.g., by using
degenerate PCR primers or heterologous DNA sequence
probes) and coexpressed.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Carnobacterium psilium</em></td>
<td>ATCC19096</td>
</tr>
<tr>
<td><em>Alcylcobacteriaceae acidothermi</em></td>
<td>ATCC40925</td>
</tr>
<tr>
<td><em>Alcylcobacteriaceae acidocaldarus</em></td>
<td>ATCC27099</td>
</tr>
</tbody>
</table>

*Uses cyclohexyl/phenacyl-CoA and not cyclohexyl/phenacyl-CoA as precursor for cycF biosynthesis.

Controlling Degree of Saturation in Olefins

The degree of saturation in olefins can be con-
trolled by regulating the degree of saturation of olefin
intermediates, for example, fatty acids. The sfa, gas, and fab
families of genes can be expressed or overexpressed to
control the saturation of fatty acids. For example, host cells
can be engineered to produce unsaturated fatty acids by
engineering the host cells to overexpress fabB or by growing
the host cells at low temperatures (e.g., less than 37°C).
FabB has preference for cis-8-decenoyl-ACP and results in
unsaturated fatty acid production in *E. coli*. Overexpression
of fabB results in the production of a percentage of
unsaturated fatty acids (de Mendoza et al., J. Biol. Chem., 258:2098-101, 1983). fabB may be inserted into and
expressed in host cells not naturally having the gene. These
unsaturated fatty acids can then be used as substrates in host
cells that are engineered to produce olefins.

Alternatively, a repressor of fatty acid biosynthesis,
for example, fabR (GenBank accession NP_418398), can be
deleted, which will result in increased unsaturated fatty acid
production in *E. coli* (Zhang et al., J. Biol. Chem., 277: 15558, 2002). Similar deletions may be made in other host
cells. A further increase in unsaturated fatty acids can be
achieved, for example, by overexpressing fabM (trans-2,
cis-3-decenoyl-ACP isomerase, GenBank accession
DAA05501) and controlled expression of fabK (trans-2-
enoyl-ACP reductase II, GenBank accession NP_357969)
from *Streptococcus pneumoniae* (Marrakchi et al., J. Biol. Chem. 277: 44809, 2002), while deleting *E. coli* fabI (trans-
enoyl-ACP reductase, GenBank accession NP_415804).
In some examples, the endogenous fabI gene can be attenuated.
Thus, increasing the percentage of palmitoleate (C16:1)
produced.

Genetic Engineering of Host Cells to Express Olefins

Various host cells can be used to produce olefins, as
described herein. A host cell can be a prokaryotic or
eukaryotic cell. For example, a polypeptide described herein
can be expressed in bacterial cells, such as *E. coli*, insect
cells, yeast or mammalian cells (such as Chinese hamster
ovary cells (CHO) cells, COS cells, VERO cells, BEK cells,
HeLa cells, Cv1 cells, MDCK cells, 293 cells, 3T3 cells, or
PC12 cells). Other exemplary host cells include cells from
the members of the genus *Escherichia*, *Bacillus*, *Listoba-
cillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*,
*Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizopo-
cor*, *Kluveromyces*, *Pichia*, *Mucor*, *Myceiophtora*, *Penicil-
lium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*,
*Saccharomyces*, *Schizosaccharomyces*, *Yarrowia*, or *Strep-
tomyces*. Yet other exemplary host cells can be a *Bacillus
lentus* cell, a *Bacillus brevis* cell, a *Bacillus stearother-
ophilus* cell, a *Bacillus licheniformis* cell, a *Bacillus alkalo-
philus* cell, a *Bacillus coagulans* cell, a *Bacillus circulans*
cell, a *Bacillus pumilis* cell, a *Bacillus thuringiensis* cell, a
*Bacillus clausii* cell, a *Bacillus megaterium* cell, a *Bacillus
subtilis* cell, a *Bacillus amylobacteriens* cell, a
*Trichoderma koningii* cell, a *Trichoderma viride* cell, a
*Trichoderma reesei* cell, a *Trichoderma longibrachiatum*
cell, an *Aspergillus awamori* cell, an *Aspergillus fumigates*
cell, an *Aspergillus foetidus* cell, an *Aspergillus nidulans*
cell, an *Aspergillus niger* cell, an *Aspergillus oryzae* cell, a
*Humicola insolens* cell, a *Humicola lanuginose* cell, a
*Rhizomucor miehei* cell, a *Mucor miehei* cell, a *Streptomycoses
lividans* cell, a *Streptomyces murinus* cell, or an Actinomy-
cetes cell. In a preferred embodiment, the host cell is an *E.
coli* cell. In a more preferred embodiment, the host cell is
from *E. coli* strains B, C, K, or W. Other suitable host cells
are known to those skilled in the art.

Various methods well known in the art can be used to
genetically engineer host cells to produce olefins. The
methods include the use of vectors, preferably expression
vectors, containing a nucleic acid encoding an ORF880
polypeptide, polypeptide variant, or a fragment thereof.
As used herein, the term “vector” refers to a nucleic acid
molecule capable of transporting another nucleic acid to
which it has been linked. One type of vector is a “plasmid,”
which refers to a circular double stranded DNA loop into
which additional DNA segments can be ligated. Another
way of a vector is a viral vector, wherein additional DNA
segments can be ligated into the viral genome. Certain
vectors are capable of autonomous replication in a host cell
into which they are introduced (e.g., bacterial vectors having
a bacterial origin of replication and episomal mammalian
vectors). Other vectors (e.g., non-episomal mammalian vec-
tors) are integrated into the genome of a host cell upon
introduction into the host cell and are thereby replicated
along with the host genome. Moreover, certain vectors, such
as expression vectors, are capable of directing the expression
of genes to which they are operatively linked. In general,
expression vectors used in recombinant DNA techniques are
often in the form of plasmids. However, other forms of
expression vectors, such as viral vectors (e.g., replication
defective retroviruses, adenoviruses, and adeno-associated
viruses), can also be used.

The recombinant expression vectors described
herein include a nucleic acid described herein in a form
suitable for expression of the nucleic acid in a host cell. The
recombinant expression vectors can include one or more
control sequences, selected on the basis of the host cell to be
used for expression. The control sequence is operably linked
to the nucleic acid sequence to be expressed. Such control
sequences are described, for example, in Goeddel, Gene
Expression Technology: Methods in Enzymology 185, Aca-
include those that direct constitutive expression of a nucleo-
tide sequence in many types of host cells and those that
direct expression of the nucleotide sequence only in certain
host cells (e.g., tissue-specific regulatory sequences). It will
be appreciated by those skilled in the art that the design of
the expression vector can depend on such factors as the
choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors described herein can be introduced into host cells to produce polypeptides, including fusion polypeptides, encoded by the nucleic acids as described herein.

[0208] Recombinant expression vectors can be designed for expression of an ORF880 polypeptide or variant in prokaryotic or eukaryotic cells (e.g., bacterial cells, such as E. coli, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells). Suitable host cells are discussed further in Goeppel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example, by using T7 promoter regulatory sequences and T7 polymerase.

[0209] Expression of polypeptides in prokaryotes, for example, E. coli, is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: (1) to increase expression of the recombinant polypeptide; (2) to increase the solubility of the recombinant polypeptide; and (3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide. This enables separation of the recombinant polypeptide from the fusion moiety after purification of the fusion polypeptide. Examples of such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin, and enterokinase. Exemplary fusion expression vectors include pGEX (Pharmacia Biotech Inc.; Smith et al., Gene (1988) 67:31-40), pMAL (New England Biolabs, Beverly, Mass.), and pPRTS (Pharmacia, Piscataway, N.J.), which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant polypeptide.

[0210] Examples of inductive, non-fusion E. coli expression vectors include pLrc (Amann et al., Gene (1988) 69:301-315) and pET 1ld (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). Target gene expression from the pLrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 1ld vector relies on transcription from a T7 gna10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gna1). This viral polymerase is supplied by host strains BL21(DE3) or HMS 174(DE3) from a resident λ prophage harboring a T7 gna1 gene under the transcriptional control of the lacUV 5 promoter.

[0211] One strategy to maximize recombinant polypeptide expression is to express the polypeptide in a host cell with an impaired capacity to proteolytically cleave the recombinant polypeptide (see Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Another strategy is to alter the nucleic acid sequence to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the host cell (Wada et al., Nucleic Acids Res. (1992) 20:2111-2118). Such alteration of nucleic acid sequences can be carried out by standard DNA synthesis techniques.

[0212] In another embodiment, the host cell is a yeast cell. In this embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast S. cerevisiae include pYEpSec1 (Balldal et al., EMBO J. (1987) 6:229-234), pMFA (Kurjan et al., Cell (1982) 30:933-945), pJRY88 (Schultz et al., Gene (1987) 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and  (Invitrogen Corp., San Diego, Calif.).

[0213] Alternatively, a polypeptide described herein can be expressed in insect cells using baculovirus expression vectors. Baculoviruses available for expression of proteins in cultured insect cells (e.g., SF9 cells) include, for example, the pAc series (Smith et al., Mol. Cell Biol. (1983) 3:2156-2165) and the pVL series (Lucklow et al., Virology (1989) 170:31-39).

[0214] In yet another embodiment, the nucleic acids described herein can be expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, Nature (1987) 329:840) and pMT2PC (Kaufman et al., EMBO J. (1987) 6:187-195). When used in mammalian cells, the expression vector’s control functions can be provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus, and Simian Virus 40. Other suitable expression systems for both prokaryotic and eukaryotic cells are described in chapters 16 and 17 of Sambrook et al., eds., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0215] Vectors can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection” refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in, for example, Sambrook et al. (supra).

[0216] For stable transformation of bacterial cells, it is known that, depending upon the expression vector and transformation technique used, only a small fraction of cells will take-up and replicate the expression vector. In order to identify and select these transformants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Selectable markers include those that confer resistance to drugs, such as ampicillin, kanamycin, chloramphenicol, or tetracycline. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0217] For stable transformation of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a...
selectable marker (e.g., resistance to antibiotics) can be introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin, and methotrexate. Nucleic acids encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide described herein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., G418 resistance). The selectable marker gene will survive, while the other cells die.

[0218] In certain instances, an ORF880 polypeptide, polypeptide variant, or a fragment thereof, is produced in a host cell that contains a naturally occurring mutation that results in an increased level of fatty acids in the host cells. In other instances, the host cell is genetically engineered to increase the level of fatty acids in the host cell relative to a corresponding wild-type host cell. For example, the host cell can be genetically engineered to express a reduced level of an acyl-CoA synthase relative to a corresponding wild-type host cell. In one embodiment, the level of expression of one or more genes, e.g., an acyl-CoA synthase gene, is reduced by genetically engineering a "knock out" host cell.

[0219] "Gene knockout" refers to a procedure by which the gene encoding the target protein is modified or inactivated so to reduce or eliminate the function of the intact protein. Inactivation of the gene may be performed by general methods such as mutagenesis by UV irradiation or treatment with N-methyl-N-nitro-N-nitrosooguanidine, site-directed mutagenesis, homologous recombination, insertion-deletion mutagenesis, or "Red-driven integration" (Datsenko et al., Proc. Natl. Acad. Sci. USA, 97:6640-45, 2000). For example, in one embodiment, a construct is introduced into a host cell, such that it is possible to select for homologous recombination events in the host cell. One of skill in the art can readily design a knock-out construct including both positive and negative selection signals for efficiently selecting transfected cells that undergo a homologous recombination event with the construct. The alteration in the host cell may be obtained, e.g., by replacing through a single or double crossover recombination a wild type DNA sequence by a DNA sequence containing the alteration. For convenient selection of transformants, the alteration may, e.g., be a DNA sequence encoding an antibiotic resistance marker or a gene complementing a possible auxotrophy of the host cell. Mutations include, but are not limited to, deletion-insertion mutations. An example of such an alteration includes a gene disruption, i.e., a perturbation of a gene such that the product is normally produced from this gene is not produced in a functional form. This could be due to a complete deletion, a deletion and insertion of a selective marker, an insertion of a selective marker, a frameshift mutation, an in-frame deletion, or a point mutation that leads to premature termination. In some instances, the entire mRNA for the gene is absent. In other situations, the amount of mRNA produced varies.

[0220] Any known acyl-CoA synthase gene can be reduced or knocked out in a host cell. Non-limiting examples of acyl-CoA synthase genes include fadD, fadK, BHI3103, yhfL, PII-4354, EAV15023, fadD1, fadD2, RPG-4074, fadD35, fadD22, faa3p or the gene encoding the protein ZP_01644857. Specific examples of acyl-CoA synthase genes include fadD35 from M. tuberculosis H37Rv [NP_217464], fadD from E. coli [YP_416216], fadD from Acinetobacter sp. ADP1 [YP_045024], fadD from Haemophilus influenzae RdK2W0 [NP_438551], fadD from Rhodopseudomonas palustris Bis B18 [YP_533919], BH3101 from Bacillus halodurans C-125 [NP_243969], P1-4354 from Pseudomonas fluorescens Pto-1 [YP_350082], EAV15023 from Comamonas testosterone KF-1 [ZP_01520072], yhfL from B. subtilis [NP_388908], fadD1 from P. aeruginosa PA01 [NP_251989], fadD1 from Ralstonia solanacearum GMI1000 [NP_520978], fadD2 from P. aeruginosa PA01 [NP_251990], the gene encoding the protein ZP_01644857 from Stenotrophomonas maltophilia R551-3, faa3p from Saccharomyces cerevisiae [NP_012257], faa1p from Saccharomyces cerevisiae [NP_014962], lefA from Bacillus subtilis [CA99571], or those described in Shockey et al., Plant. Physiol. 129: 1710-1722, 2002; Caviglia et al., J. Biol. Chem. 279: 1163-1169, 2004; Knoll et al., J. Biol. Chem. 269(23):16348-56, 1994; Johnson et al., J. Biol. Chem. 269: 18037-18046, 1994; and Black et al., J. Biol. Chem. 267: 25513-25520, 1992.

Transport Proteins

[0221] Transport proteins can export polypeptides and hydrocarbons (e.g., olefins) out of a host cell. Many transport and efflux proteins serve to excrete a wide variety of compounds and can be naturally modified to be selective for particular types of hydrocarbons.

[0222] Non-limiting examples of suitable transport proteins are ATP-Binding Cassette (ABC) transport proteins, efflux proteins, and fatty acid transporter proteins (FATP). Additional non-limiting examples of suitable transport proteins include the ABC transport proteins from organisms such as Caenorhabditis elegans, Arabidopsis thaliana, Alka- lignes eutrophus, and Rhodococcus erythropolis. Exam- plary ABC transport proteins that can be used are listed in FIG. 10 (e.g., CER5, AtMRP5, AmiS2, and AtGP1). Host cells can also be chosen for their endogenous ability to secrete hydrocarbons. The efficiency of hydrocarbon production and secretion into the host cell environment (e.g., culture medium, fermentation broth) can be expressed as a ratio of intracellular product to extracellular product. In some examples, the ratio can be about 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4, or 1:5.

Fermentation

[0223] The production and isolation of olefins can be enhanced by employing beneficial fermentation techniques. One method for maximizing production while reducing costs is increasing the percentage of the carbon source that is converted to hydrocarbon products.

[0224] During normal cellular lifecycles, carbon is used in cellular functions, such as producing lipids, saccharides, proteins, organic acids, and nucleic acids. Reducing the amount of carbon necessary for growth-related activities can increase the efficiency of carbon source conversion to product. This can be achieved by, for example, first growing host cells to a desired density (for example, a density achieved at the peak of the log phase of growth). At such a point, replication checkpoint genes can be harnessed to stop the growth of cells. Specifically, quorum sensing mechanisms (reviewed in Camilli et al., Science 311:1113, 2006; Venturi FEMS Microbiol. Rev. 50:274-291, 2006; and Reading et al.,
Genes that can be activated to stop cell replication and growth in *E. coli* include umuDC genes. The overexpression of umuDC genes stops the progression from stationary phase to exponential growth (Murli et al., *J. of Bact.* 182:1127, 2000). UmuC is a DNA polymerase that can carry out translesion synthesis over non-coding lesions—the mechanistic basis of most UV and chemical mutagenesis. The umuDC gene products are involved in the process of translesion synthesis and also serve as a DNA sequence damage checkpoint. The umuDC gene products include UmuC, UmuD', UmuD''C, UmuD'x, and UmuDx. Simultaneously, product-producing genes can be activated, thus minimizing the need for replication and maintenance pathways to be used while an olefin is being made.

The percentage of input carbons converted to olefins can be a cost driver. The more efficient the process is (i.e., the higher the percentage of input carbons converted to olefins), the less expensive the process will be. For oxygen-containing carbon sources (e.g., glucose and other carbohydrate-based sources), the oxygen must be released in the form of carbon dioxide. For every 2 oxygen atoms released, a carbon atom is also released leading to a maximal theoretical metabolic efficiency of approximately 34% (w/w) (for fatty acid derived products). This figure, however, changes for other hydrocarbon products and carbon sources. Typical efficiencies in the literature are approximately less than 5%. Host cells engineered to produce olefins can have greater than about 1, 3, 5, 10, 15, 20, 25, and 30% efficiency. In one example, host cells can exhibit an efficiency of about 10% to about 25%. In other examples, such host cells can exhibit an efficiency of about 25% to about 30%. In other examples, host cells can exhibit greater than 30% efficiency.

The host cell can be additionally engineered to express recombinant cellulolytic enzymes, such as those described in PCT application number PCT/US2007/003736. These cellulolytic enzymes can allow the host cell to use cellulotic material as a carbon source. For example, the host cell can be additionally engineered to express invertases (EC 3.2.1.26) so that sucrose can be used as a carbon source. Similarly, the host cell can be engineered using the teachings described in U.S. Pat. Nos. 5,432,846; 5,482,846; 5,470,936; 5,451,539; 5,424,202; 5,424,202; 5,482,846; and 5,602,030; so that the host cell can assimilate carbon efficiently and use cellulotic materials as carbon sources.

In one example, the fermentation chamber can enclose a fermentation that is undergoing a continuous reduction. In this instance, a stable reductive environment can be created. The electron balance can be maintained by the release of carbon dioxide (in gaseous form). Efforts to augment the NADH/NADPH balance can also facilitate in stabilizing the electron balance. The availability of intracellular NADPH can also be enhanced by engineering the host cell to express an NADH:NADPH transhydrogenase. The expression of one or more NADH:NADPH transhydrogenases converts the NADH produced in glycolysis to NADPH, which can enhance the production of olefins.

For small scale production, the engineered host cells can be grown in batches of, for example, around 100 mL, 500 mL, 1 L, 2 L, 5 L, or 10 L; fermented; and induced to express desired olefins based on the specific genes encoded in the appropriate plasmids. For example, *E. coli* BL21 (DE3) cells harboring pBAD24 (with ampicillin resistance and the olefin synthesis pathway) as well as pUMVC1 (with kanamycin resistance and the acetyl-CoA/malonyl-CoA overexpression system) can be incubated overnight in 2 L flasks at 37°C shaken at >200 rpm in 500 mL LB medium supplemented with 75 μg/mL ampicillin and 50 μg/mL kanamycin until cultures reach an OD<sub>600</sub> of >0.8. Upon achieving an OD<sub>600</sub> of >0.8, the cells can be supplemented with 25 mM sodium propionate (pH 8.0) to activate the engineered gene systems for production and to stop cellular proliferation by activating UmuC and UmuD proteins. Induction can be performed for 6 h at 30°C. After incubation, the media can be examined for olefins using GC-MS.

For large scale production, the engineered host cells can be grown in batches of 10 L, 100 L, 1000 L or larger; fermented; and induced to express desired olefins based on the specific genes encoded in the appropriate plasmids. For example, *E. coli* BL21(DE3) cells harboring pBAD24 (with ampicillin resistance and the olefin synthesis pathway) as well as pUMVC1 (with kanamycin resistance and the acetyl-CoA/malonyl-CoA overexpression system) can be incubated from a 500 mL seed culture for 10 L fermentations (5 L for 100 L fermentations, etc.) in LB media (glycerol free) with 50 μg/mL kanamycin and 75 μg/mL ampicillin at 37°C, and shaken at >200 rpm until cultures reach an OD<sub>600</sub> of >0.8 (typically 16 h). Media can be continuously supplemented to maintain 25 mM sodium propionate (pH 8.0) to activate the engineered gene systems for production and to stop cellular proliferation by activating UmuC and UmuD proteins. Media can be continuously supplemented with glucose to maintain a concentration 25 g/100 mL.

After the first hour of induction, aliquots of no more than 10% of the total cell volume can be removed each hour and allowed to sit without agitation to allow the olefins to rise to the surface and undergo a spontaneous phase separation. The olefin component can then be collected, and the aqueous phase returned to the reaction chamber. The reaction chamber can be operated continuously. When the OD<sub>600</sub> drops below 0.6, the cells can be replaced with a new batch grown from a seed culture.

**Producing Olefins Using Cell-Free Methods**

In some methods described herein, an olefin can be produced using a purified polypeptide described herein and a fatty acid substrate. For example, a host cell can be engineered to express a polypeptide (e.g., an ORF880 polypeptide or variant) as described herein. The host cell can be cultured under conditions suitable to allow expression of the polypeptide. Cell free extracts can then be generated using known methods. For example, the host cells can be lysed using detergents and/or sonication. The expressed polypeptides can be purified using known methods. After obtaining the cell free extracts, fatty acid substrates described herein can be added to the cell free extracts and maintained under conditions to allow conversion of the fatty acid substrates to olefins. The olefins can then be separated and purified using known techniques.

**Post-Production Processing**

The olefins produced during fermentation can be separated from the fermentation media. Any known technique for separating olefins from aqueous media can be
used. One exemplary separation process is a two phase (bi-phasic) separation process. This process involves fermenting the genetically engineered host cells under conditions sufficient to produce an olefin, allowing the olefin to collect in an organic phase, and separating the organic phase from the aqueous fermentation broth. This method can be practiced in both a batch and continuous fermentation setting.

Bi-phasic separation uses the relative immiscibility of olefins to facilitate separation. Immiscible refers to the relative inability of a compound to dissolve in water and is defined by the compound’s partition coefficient. One of ordinary skill in the art will appreciate that by choosing a fermentation broth and organic phase, such that the olefin being produced has a high log P value, the olefin can separate into the organic phase, even at very low concentrations, in the fermentation vessel.

The olefins produced by the methods described herein can be relatively immiscible in the fermentation broth, as well as in the cytoplasm. Therefore, the olefin can collect in an organic phase either intracellularly or extracellularly. The collection of the products in the organic phase can lessen the impact of the olefin on cellular function and can allow the host cell to produce more product.

The methods described herein can result in the production of homogeneous compounds wherein at least about 60%, 70%, 80%, 90%, or 95% of the olefins produced will have carbon chain lengths that vary by less than about 6 carbons, less than about 4 carbons, or less than about 2 carbons. These compounds can also be produced with a relatively uniform degree of saturation. For example, at least about 60%, 70%, 80%, 90%, or 95% of the olefins will be monounsaturated, disaturated, or trisaturated. These compounds can be used directly as fuels, fuel additives, starting materials for production of other chemical compounds (e.g., polymers, surfactants, plastics, textiles, solvents, adhesives, etc.), or personal care additives. These compounds can also be used as feedstock for subsequent reactions, for example, hydrogenation, catalytic cracking (via hydrogenation, pyrolysis, or both), or epoxidation reactions to make other products.

In some embodiments, the olefins produced using methods described herein can contain between about 50% and about 90% carbon; between about 5% and about 25% hydrogen; or between about 5% and about 25% oxygen. In other embodiments, the olefins produced using methods described herein can contain between about 65% and about 85% carbon; between about 10% and about 15% hydrogen; or between about 10% and about 20% oxygen.

Fuel Compositions

The olefins described herein can be used as or converted into a fuel. One of ordinary skill in the art will appreciate that, depending upon the intended purpose of the fuel, different olefins can be produced and used. For example, a branched olefin may be desirable for automobile fuel that is intended to be used in cold climates. In addition, when the olefins described herein are used as a feedstock for fuel production, one of ordinary skill in the art will appreciate that the characteristics of the olefin feedstock will affect the characteristics of the fuel produced. Hence, the characteristics of the fuel product can be selected for by producing particular olefins for use as a feedstock.

Using the methods described herein, biofuels having desired fuel qualities can be produced from olefins. Biologically produced olefins represent a new source of biofuels, which can be used as jet fuel, diesel, or gasoline. Some biofuels made using olefins have not been produced from renewable sources and are new compositions of matter. These new fuels can be distinguished from fuels derived from petrochemical carbon on the basis of dual carbon-isotopic fingerprinting. Additionally, the specific source of biosourced carbon (e.g., glucose vs. glycerol) can be determined by dual carbon-isotopic fingerprinting (see, e.g., U.S. Pat. No. 7,169,588, in particular col. 4, line 31, to col. 6, line 8).

Hydrocarbons comprising biologically produced hydrocarbons, particularly α-olefins biologically produced using the fatty acid bio synthetic pathway, have not been produced from renewable sources and, as such, are new compositions of matter. These new hydrocarbons can be distinguished from hydrocarbons derived from petrochemical carbon on the basis of dual carbon-isotopic fingerprinting or 14C dating. Additionally, the specific source of biosourced carbon (e.g., glucose vs. glycerol) can be determined by dual carbon-isotopic fingerprinting (see U.S. Pat. No. 7,169,588, which is herein incorporated by reference).

The ability to distinguish biologically produced hydrocarbons from petroleum based hydrocarbons is beneficial in tracking these materials in commerce. For example, hydrocarbons or chemicals comprising both biologically based and petroleum based carbon isotope profiles may be distinguished from hydrocarbons and chemicals made only of petroleum based materials. Hence, the instant materials may be followed in commerce on the basis of their unique carbon isotope profile.

Hydrocarbons can be distinguished from petroleum based fuels by comparing the stable carbon isotope ratio (13C/12C) in each fuel. The 13C/12C ratio in a given biologically based material is a consequence of the 13C/12C ratio in atmospheric carbon dioxide at the time the carbon dioxide is fixed. It also reflects the precise metabolic pathway. Regional variations also occur. Petroleum, C3 plants (the broadleaf), C4 plants (the grasses), and marine carbonates all show significant differences in 13C/12C and the corresponding 813C values. Furthermore, lipid matter of C3 and C4 plants analyze differently than materials derived from the carbohydrate components of the same plants as a consequence of the metabolic pathway.

Within the precision of measurement, 13C shows large variations due to isotopic fractionation effects, the most significant of which for biologically produced hydrocarbons is the photosynthetic mechanism. The major cause of differences in the carbon isotope ratio in plants is closely associated with differences in the pathway of photosynthetic carbon metabolism in the plants, particularly the reaction occurring during the primary carboxylation (i.e., the initial fixation of atmospheric CO2). Two large classes of vegetation are those that incorporate the “C3” (or Calvin-Benson) photosynthetic cycle and those that incorporate the “C4” (or Hatch-Slack) photosynthetic cycle.

In C3 plants, the primary CO2 fixation or carboxylation reaction involves the enzyme ribulose-1,5-diphosphate carboxylase and the first stable product is a 3-carbon compound. C4 plants, such as hardwoods and conifers, are dominant in the temperate climate zones.
In C₄ plants, an additional carboxylation reaction involving another enzyme, phosphoenol-pyruvate carboxylase, is the primary carboxylation reaction. The first stable carbon compound is a 4-carbon acid which is subsequently decarboxylated. The CO₂ thus released is fixed by the C₃ cycle. Examples of C₄ plants are tropical grasses, corn, and sugar cane.

Both C₄ and C₃ plants exhibit a range of ¹³C/¹²C isotopic ratios, but typical values are about −7 to about −13 per mil for C₄ plants and about −19 to about −27 per mil for C₃ plants (see, e.g., Stinner et al., *Radiocarbon*, 19: 355 (1977)). Coal and petroleum fall generally in this latter range. The ¹³C measurement scale was originally defined by a zero set by Pee Dee Belemnite (PDB) limestone, where values are given in parts per thousand deviations from this material. The ¹³C values, are in parts per thousand (per mil), abbreviated ‰ and are calculated as follows:

\[
\Delta^{13}C = \text{(sample)}-\text{(standard)} / \text{(standard)} \times 1000
\]

Since the PDB reference material (RM) has been exhausted, a series of alternative RMs have been developed in cooperation with the IAEA, USGS, NIST, and other selected international isotope laboratories. Notations for the per mil deviations from PDB is δ¹³C. Measurements are made on CO₂ by high precision stable ratio mass spectrometry (IRMS) on molecular ions of masses 44, 45, and 46.

The invention provides a hydrocarbon or biofuel produced by any of the methods disclosed herein. Specifically, the hydrocarbon or biofuel can have a δ¹³C of about −28 or greater, about −27 or greater, −20 or greater, −18 or greater, −15 or greater, −13 or greater, −10 or greater, or −8 or greater. For example, the hydrocarbon can have a δ¹³C of about −30 to about −15, about −27 to about −19, about −25 to about −21, about −15 to about −5, about −13 to about −7, about −13 to about −10. In some examples, a biofuel composition can be made that includes an olefin having a δ¹³C of from about −10.9 to about −15.4, where the olefin accounts for at least about 85% of biosourced material (i.e., derived from a renewable source, such as biomass, cellulosic materials, and sugars) in the composition. The invention also provides for a hydrocarbon or biofuel with a δ¹³C of about −10, −11, −12, or −13.

Biologically produced hydrocarbons can also be distinguished from petroleum based hydrocarbons by comparing the amount of ¹³C in each fuel. Because ¹³C has a nuclear half life of 5730 years, petroleum based fuels containing “older” carbon can be distinguished from biofuels which contain “newer” carbon (see, e.g., Carrié, “Source Apportionment of Atmospheric Particles,” *Characterization of Environmental Particles*, J. Buffle and H. P. van Leeuwen, Eds., 1 of Vol. I of the IUPAC Environmental Analytical Chemistry Series (Lewis Publishers, Inc.) (1992) 3-74).

The basic assumption in radiocarbon dating is that the constancy of ¹³C concentration in the atmosphere leads to the constancy of ¹³C in living organisms. However, because of atmospheric nuclear testing since 1950 and the burning of fossil fuel since 1850, ¹³C has acquired a second, geochemical time characteristic. Its concentration in atmospheric CO₂, and hence in the living biosphere, approximately doubled at the peak of nuclear testing, in the mid-1960s. It has since been gradually returning to the steady-state cosmogenic (atmospheric) baseline isotope rate (¹³C/¹²C) of about 1.2×10⁻¹², with an approximate relaxation “half-life” of 7-10 years. (This latter half-life must not be taken literally; rather, one must use the detailed atmospheric nuclear input/decay function to trace the variation of atmospheric and biospheric ¹³C since the onset of the nuclear age.)

It is this latter biospheric ¹³C time characteristic that holds out the promise of annual dating of recent biospheric carbon. ¹³C can be measured by accelerator mass spectrometry (AMS), with results given in units of “fraction of modern carbon” (Fm). Fm is defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C. As used herein, “fraction of modern carbon” or “Fm” has the same meaning as defined by National Institute of Standards and Technology (NIST) Standard Reference Materials (SRMs) 4990B and 4990C, known as oxalic acids standards HOx1 and HOxII, respectively. The fundamental definition relates to 0.95 times the ¹³C/¹²C isotope ratio HOx1 (referred to AD 1950). This is roughly equivalent to decay-corrected pre-Industrial Revolution wood. For the current living biosphere (plant material), Fm is approximately 1.1.

The invention provides a hydrocarbon or biofuel which can have an Fm of at least about 1. For example, the hydrocarbon or biofuel can have an Fm of at least about 1.01, an Fm of about 1 to about 1.5, an Fm of about 1.04 to about 1.18, or an Fm of about 1.11 to about 1.24. In some examples, the olefin in the biofuel composition can have a fraction of modern carbon (Fm of) of, for example, at least about 1.003, 1.010, or 1.5.

Another measurement of ¹³C is known as the percent of modern carbon, pMC. For an archaeologist or geologist using ¹³C dates, AD 1950 equals “zero years old”. This also represents 100 pMC. “Bomb carbon” in the atmosphere reached almost twice the normal level in 1963 at the peak of thermo-nuclear weapons. Its distribution within the atmosphere has been approximated since its appearance, showing values that are greater than 100 pMC for plants and animals living since AD 1950. It has gradually decreased over time with today’s value being near 107.5 pMC. This means that a fresh biomass material, such as corn, would give a ¹³C signature near 107.5 pMC. Petroleum based compounds will have a pMC value of zero. Combining fossil carbon with present day carbon will result in a dilution of the present day pMC content. By presuming 107.5 pMC represents the ¹³C content of present day biomass materials and 0 pMC represents the ¹³C content of petroleum based products, the measured pMC value for that material will reflect the proportions of the two component types. For example, a material derived 100% from present day soybeans would give a radiocarbon signature near 107.5 pMC. If that material was diluted 50% with petroleum based products, it would give a radiocarbon signature of approximately 54 pMC.

A biologically based carbon content is derived by assigning 100% equal to 107.5 pMC and 0% equal to 0 pMC. For example, a sample measuring 99 pMC will give an equivalent biologically based carbon content of 93%. This value is referred to as the mean biologically based carbon result and assumes all the components within the analyzed material originated either from present day biological material or petroleum based material.

The invention provides a hydrocarbon or biofuel which can have a pMC of at least about 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, or 100. The invention further provides for a hydrocarbon or fuel which has a pMC of
between about 50 to about 100, about 60 to about 100, about 70 to about 100, about 80 to about 100, about 85 to about 100, and about 87 to about 98, about 90 to about 95. The invention further provides for a hydrocarbon or biofuel with a pMC of about 50, 91, 92, 93, 94, or 94.2.

[0256] Fuel additives are used to enhance the performance of a fuel or engine. For example, fuel additives can be used to alter the freezing/gelling point, cloud point, lubricity, viscosity, oxidative stability, ignition quality, octave level, and/or flash point. In the United States, all fuel additives must be registered with the Environmental Protection Agency. The names of fuel additives and the companies that sell the fuel additives are publicly available by contacting the EPA or by viewing the agency’s website. One of ordinary skill in the art will appreciate that the olefin-based biofuels described herein can be mixed with one or more fuel additives to impart a desired quality.

[0257] The olefin-based biofuels described herein can be mixed with other fuels, such as various alcohols, such as ethanol and butanol, and petroleum-derived products, such as gasoline, diesel, or jet fuel.

[0258] In some examples, the mixture can include at least about 10%, 15%, 20%, 30%, 40%, 50%, or 60% by weight of the olefin. In other examples, a biofuel composition can be made that includes at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or 95% of an olefin that includes a carbon chain that is 8:1, 10:1, 12:1, 13:1, 14:1, 14:2, 15:1, 15:2, 16:1, 16:2, 17:1, 17:2, 18:1, 18:2, 18:3, 19:1, 19:2, 19:3, 20:1, 20:2, 20:3, 22:1, 22:2, or 22:3. Such biofuel compositions can additionally include at least one additive selected from a cloud point lowering additive that can lower the cloud point to less than about 5°C, or 0°C; a surfactant; a microemulsion; at least about 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, or 95% diesel fuel from triglycerides; petroleum-derived gasoline; or diesel fuel from petroleum.

EXAMPLES

[0259] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Example 1

Identification and Reclassification of a Microorganism Belonging to the Genus Jeotgalicoccus that is an α-Olefin Producer

[0260] Micrococcus candidus ATCC 8456 was previously reported to synthesize aliphatic hydrocarbons with carbon chains lengths ranging from C₁₈ to C₂₀ (Morrison et al., J. Bacteriol. 108:353-358, 1971). To identify the hydrocarbons produced by this strain, ATCC 8456 cells were cultured in 15 mL TSBYE medium (3% Tryptic Soy Broth+0.5% Yeast Extract), for 40-48 h at 30°C. Cells from 5 mL of culture were pelleted, resuspended in 1 mL methanol, sonicated for 30 min and extracted with 4 mL hexane. After solvent evaporation, samples were resuspended in 0.1 mL hexane and analyzed by GC-MS. The hydrocarbons were identified as the following α-olefins: 15-methyl-1-heptadecene (α-C₁₇), 16-methyl-1-heptadecene (i-C₁₇), 1-nonadecene (α-C₁₉), 17-methyl-1-nonadecene (α-C₁₉), and 18-methyl-1-nonadecene (i-C₁₀) (i-iso, α-anteiso, n-straight chain) (see, e.g., FIG. 1 and FIG. 2).

[0261] Based upon the following analyses, it was determined that ATCC 8456 was previously misidentified as belonging to the genus Micrococcus. The phylogenetic classification of ATCC 8456 was reassessed by amplifying and sequencing the partial 16s rRNA gene using primers Eub338f and 1492R (see DeLong et al., PNAS 89:5685, 1992). The 16s rRNA sequence of ATCC8456 (SEQ ID NO:3, depicted in FIG. 7C) was analyzed using the classifier program of the Ribosomal Database Project II. Based upon this analysis, the strain was identified as belonging to the genus Jeotgalicoccus. The genus Jeotgalicoccus has been previously described (see, e.g., Jung-Elmo et al., Int. J. Syst. Evol. Microbiol. 53:595-602, 2003).

[0262] Additional analysis using the G+C content of ATCC 8456 was conducted. Jeotgalicoccus is a low G+C Gram-positive bacteria related to the genus Staphylococcus (see FIG. 3). Micrococcus are high G+C Gram-positive bacteria. The ends of several clones from a cosmid library of ATCC 8456 genomic DNA were sequenced. Based upon a DNA sequence of approximately 4,000 bp, the G+C content was determined to be approximately 36%. Nucleotide sequence searches against a non-redundant protein database revealed that all sequences with a match to a database entry were similar to proteins from low G+C Gram-positive bacteria, such as species belonging to the genus Staphylococcus or Bacillus, but not the genus Micrococcus.

[0263] Next, an analysis of the entire genome of ATCC 8456 was conducted. Based on a DNA sequence of approximately 2.1 MB, the G+C content of the entire genome was determined to be about 36.7%. In contrast, bacteria of the genus Micrococcus are known to have high G+C genomes. For example, the genome of Micrococcus luteus NCTC 2665 has a G+C content of 72.9% (GenBank Accession ABLQ01000001-68). Based upon the G+C content analysis, it was determined that the ATCC 8456 microorganism does not belong to the genus Micrococcus.

[0264] Additional Jeotgalicoccus strains were also examined to determine if they produced α-olefins. The following strains of Jeotgalicoccus were examined: Jeotgalicoccus halotolerans DSMZ 17274, Jeotgalicoccus psychrophilus DSMZ 19085, and Jeotgalicoccus pinnipedialis DSMZ 17030. Each strain was cultured in 15 mL TSBYE medium (3% Tryptic Soy Broth+0.5% Yeast Extract), and the hydrocarbons were isolated and analyzed by GC-MS as described above. All three strains produced α-olefins similar to the ones produced by ATCC 8456 (FIGS. 1B, 1C, and 1D depict GC-MS traces for the hydrocarbons produced by Jeotgalicoccus halotolerans DSMZ 17274 cells, Jeotgalicoccus pinnipedialis DSMZ 17030 cells, and Jeotgalicoccus psychrophilus DSMZ 19085 cells, respectively). These data indicate that the ability to produce α-olefins is widespread among the genus Jeotgalicoccus.

Example 2

Production of Increased Levels of Olefins and α-Olefinss Not Normally Produced by ATCC 8456 Cells Using Fatty Acid Feeding

[0265] The fatty acids eicosanoic acid (straight-chain C₂₀ fatty acid), 16-methyl octadecanoic acid, and 17-methyl octadecanoic acid (branched-chain C₁₇ fatty acids) were identified as components of ATCC 8456’s lipids. These fatty acids were deduced to be the direct precursors, after decarboxylation, for 1-nonadecene, 15-methyl-1-heptadecene,
and 16-methyl-1-heptadecene biosynthesis, respectively. In order to improve \( \alpha \)-olefin production and to produce olefins not normally produced by ATCC 8456 cells, fatty acid feeding experiments were carried out as described below. ATCC 8456 cells were grown in 15 mL TSBYE medium (3% Tryptic Soy Broth+0.5% Yeast Extract). Fatty acids were added to the culture medium at final concentrations of 0.5 g/L (0.05%). After growth for 40-48 h at 30°C, cells from 5 mL of culture were pelleted, resuspended in 1 mL methanol, sonicated for 30 min, and extracted with 4 mL hexane. After solvent evaporation, samples were resuspended in 0.1 mL hexane and analyzed by GC/MS.

When cultures were fed eicosanoic acid, an increase in 1-nondecane production of approximately 18-fold was observed (see Fig. 4A; black traces depict production without fatty acid feeding and gray traces depict production with fatty acid feeding). When cultures were fed stearic acid or palmitic acid, an increase in the production of the \( \alpha \)-olefins 1-pentadecene and 1-heptadecene, respectively, was observed (see Fig. 4B). These \( \alpha \)-olefins are not normally produced by ATCC 8456 cells. This indicated that fatty acids were the direct precursors for \( \alpha \)-olefins and that *Jeotgalicoccus* bacteria could be used to enzymatically convert fatty acids into \( \alpha \)-olefins in vivo.

In alternative methods, when resting *Jeotgalicoccus* cells were fed various fatty acids and the production of \( \alpha \)-olefins is analyzed, similar results are observed.

**Example 3**

**In Vitro Synthesis of \( \alpha \)-Olefins Using Cell Extracts and Partially Purified Proteins**

A cell free extract of ATCC 8456 was used to convert free fatty acids into \( \alpha \)-olefins. The cell free extract was generated using the following procedure: ATCC 8456 cells were grown in TSBYE medium (3% Tryptic Soy Broth+0.5% Yeast Extract) at 30°C for 24 h with shaking. The cells were then pelleted from the culture by centrifuging at 3,700 rpm for 20 min. The cell pellet was then resuspended in 50 mM Tris buffer pH 7.5 with 0.1 M NaCl and 2.0 mM dithiothreitol to a concentration of 0.1 g/mL cells. To this cell slurry, 200 units/mL of lysozyme (Sigma) were added on ice. The cell lysis reaction was allowed to proceed for 30 min. The cells were then sonicated at 12 W on ice for three cycles of 1.5 sec of sonication followed by 1.5 sec of rest. Sonication lasted for a total of 9 sec. This procedure was repeated 5 times with a 1 min interval between each sonication cycle. The lysed cells were then centrifuged at 12,000 rpm for 10 min to pellet the cell debris. The supernatant (i.e., cell free extract) was then removed and used for the conversion of free fatty acids to \( \alpha \)-olefins.

After obtaining the cell free extract, the free fatty acids, stearic acid and eicosanoic acid, were converted to \( \alpha \)-olefins using the cell free extract as described below.

First, a 5% stock solution of sodium or potassium stearate was made in 1% Tergitol solution (Sigma, St. Louis, Mo.). Next, 6 µL of the stock solution was added to 1 mL of the cell free extract at room temperature to obtain a final concentration of 1 mM free fatty acid salt. The reaction was conducted at room temperature for 3 h. The \( \alpha \)-olefins were recovered by adding 200 µL of ethyl acetate to the mixture, vortexing briefly, centrifuging briefly, and then removing the organic phase. The \( \alpha \)-olefins were detected using GC/MS.

**FIG. 5** shows the GC/MS trace for assaying stearic acid. In sample 1 (TIC1; black line in FIG. 5), no stearic acid was added to the cell free extract. In sample 2 (TIC3; blue line in FIG. 5), the cell free extract was replaced with 50 mM Tris pH 7.5 buffer with 0.1 M sodium chloride to which stearic acid was added. In sample 3 (TICS; red line in FIG. 5), stearic acid was added to the cell free extract. The peak at 7.62 min had the same retention time and the same mass spectra as a 1-heptadecene standard (Sigma). When eicosanoic acid was added under similar conditions, 1-nondecane was formed.

Boiling the cell free extract eliminated the production of \( \alpha \)-olefins upon the addition of free fatty acids. This data strongly suggested that the ATCC 8456 catalyst was protein based.

**FIG. 4** shows the GC/MS trace for assaying stearic acid. In sample 1 (TIC1; black line in FIG. 5), no stearic acid was added to the cell free extract. In sample 2 (TIC3; blue line in FIG. 5), the cell free extract was replaced with 50 mM Tris pH 7.5 buffer with 0.1 M sodium chloride to which stearic acid was added. In sample 3 (TICS; red line in FIG. 5), stearic acid was added to the cell free extract. The peak at 7.62 min had the same retention time and the same mass spectra as a 1-heptadecene standard (Sigma). When eicosanoic acid was added under similar conditions, 1-nondecane was formed.

**Example 4**

**Purification and Identification of a Protein that Converts Fatty Acids into \( \alpha \)-Olefins**

To isolate the protein necessary for \( \alpha \)-olefin production from ATCC 8456 cells, the following protein purification procedure was carried out. First, 6 L of ATCC 8456 cells were cultured in TSBYE medium at 30°C for 24 h with shaking. The cells were pelleted at 3,700 rpm for 20 min at 4°C, and the supernatant was discarded. The cell pellet was resuspended in a solution of 100 mL of 50 mM Tris pH 8.0, 0.1 M NaCl, 2.0 mM DTT, and bacterial protease inhibitors. The cell slurry was then passed through a French press one time at a pressure of 30,000 psi. Next, the cell slurry was sonicated as described in Example 3 to shear the DNA. The cell free extract was then centrifuged at 10,000 rpm for 60 min at 4°C. The supernatant was then...
removed and ammonium sulfate was added to a final concentration of 50% (wt/vol) ammonium sulfate saturation. The mixture was gently stirred at 4°C for 60 min and then centrifuged at 10,000 rpm for 30 min. The supernatant was then removed and additional ammonium sulfate was added to 65% (wt/vol) saturation. The mixture was stirred again for 60 min at 4°C and centrifuged at 10,000 rpm for 30 min. The supernatant was discarded. The remaining pellet was resuspended in 50 mL of 50 mM Tris pH 8.0 and 2.0 mM DTT.

[0277] The mixture was passed through a 5 mL HiTrap SP column (GE Healthcare) at 3 mL/min and 4°C. The following buffers were used as an elution gradient: buffer A contained 50 mM Tris pH 8.0 and 2.0 mM DTT and buffer B contained 50 mM Tris pH 8.0, 1.0 M NaCl, and 2.0 mM DTT. After the column was loaded with the mixture, the column was washed with 40% buffer B. Next a 20 min gradient of 40% buffer B to 100% buffer B at 3.0 mL/min was carried out. 5 mL fractions were collected during the elution gradient. Each fraction was tested for activity as described in Example 3. Fractions containing α-olefin production activity typically eluted between 600 and 750 mM NaCl concentration. Fractions containing activity were then pooled and dialyzed into buffer A.

[0278] The dialyzed protein fraction was then loaded onto a 1 mL ResourceQ (GE Healthcare) column at 4 mL/min at 4°C. Buffer B used with the HiTrap SP column was used for the ResourceQ column. A 7-min elution gradient between 0% buffer B and 25% buffer B was run at 4 mL/min. 1.5 mL fractions were collected and assayed for activity. Active fractions eluted between 150 and 200 mM NaCl concentrations. Fractions containing activity were then pooled and concentrated with a Millipore Amicon protein concentrator (4 mL and 10 kDa exclusion size) to about 50 µL. The approximate protein concentration was determined with a Bradford assay (Bio Rad). Final protein concentrations ranged from about 5 mg/mL to about 10 mg/mL. 30 µL of protein was then loaded onto a SDS PAGE gel (Invitrogen) along with an appropriate protein molecular weight marker. The gel was stained with Simple Safe Coomassie stain (Invitrogen). Fig. 6 depicts a representative gel. Two intense protein bands at 50 kDa and 20 kDa were observed.

[0279] To determine the identity of the protein bands, the bands were excised from the gel, digested with trypsin, and analyzed using LC/MS/MS. The LC/MS/MS data was analyzed using the program Mascot (Mann et al., Anal. Chem. 66:4390-4399, 1994). The ATCC 8456 genome was sequenced. The genomic data was used to interpret the LC/MS/MS data and to determine the identity of the protein bands. The 50 kDa band had a strong match with ORF880. The Mascot score assigned to this match was 919, a high score. Furthermore, ORF880 has a predicted molecular weight of 48,367 Da. The nucleotide and amino acid sequences of ORF880 are presented in FIGS. 7A (SEQ ID NO:1) and 7B (SEQ ID NO:2), respectively.

Example 5

Heterologous Expression of Jeotgalicoccus ATCC 8456 Orf880 in E. coli

[0280] Jeotgalicoccus ATCC 8456 ORF880 was identified as one of the two major proteins in a highly purified enzyme fraction that catalyzed the conversion of free fatty acids to α-olefins. The genomic DNA encoding ATCC 8456_orf880 was cloned into pCDF-Duet1 under the control of the T7 promoter, and E. coli was transformed with various vectors, as described below. The E. coli cells were grown, and the hydrocarbons produced by the cells were analyzed as described in Example 2. When 0.05% stearic acid was fed to cultures of E. coli transformed with the 8456_orf880-containing vector, the expression of 8456_orf880 led to the formation of 1-heptadecene in E. coli (see FIG. 8, depicting GC/MS traces of α-olefins from E. coli either without (black) or with (gray) expression of 8456_orf880). In contrast, adding 0.05% stearic acid to cultures of E. coli transformed with a vector control (not containing ATCC_orf880) did not result in the production of 1-heptadecene. This demonstrated that 8456_orf880 expressed from free fatty acids in an E. coli heterologous host. This result indicates that α-olefin biosynthesis can be performed in heterologous organisms. Additionally, when E. coli cells expressing 8456_orf880 protein were fed with 0.05% palmitic acid or 0.05% eicosanoic acid, the production of 1-pentadecene or 1-nonadecene, respectively, was observed.

Example 6

Production of α-Olefinas from Glucose by Heterologous Expression of Jeotgalicoccus ATCC 8456 orf880 in E. coli MG1655 ΔfadD

[0281] Construction of fadD Deletion Strain

[0282] The fadD gene of E. coli MG1655 was deleted using the lambda red system (Datsenko et al., 2000, Proc. Natl. Acad. Sci. USA. 97: 6640-6645) as described below.

[0283] The chloramphenicol acetyltransferase gene from pKD3 was amplified with the primers

fad1: (5′-TAACCGGGCTTCGACCACTGACTTAACGCTCAGGCTTTATTGCTTTCG-3′), and

fad2: (5′-CAATTTGCGGTAGCTGACGAGCAATCGCATTTTTAGGGTGAGAGA ATGCATATGAAATATCCCTTTGATCC-3′).
Expression of *Jeotgalicoccus* ATCC 8456 _Orf880_ in *E. coli* MG1655 ΔfadD

[0285] The genomic DNA encoding *ATCC 8456_orf880*, which was codon-optimized for expression in *E. coli* (SEQ ID NO:4), was cloned into vector OP80 (pCL1920 derivative) under the control of the _P_{lac}^n_ promoter, and *E. coli* MG1655 ΔfadD was transformed with the resulting vector. The _E. coli_ cells were grown at 37°C in M9 mineral medium supplemented with 20 μg/mL unc105 and 100 μg/mL spectinomycin. Glucose (1%, w/v) was the only source of carbon and energy. When the culture reached an OD_{600} of 0.8 to 1.0, IPTG (1 mM) and 0.5 mM delta-aminolaevulinic acid (heme biosynthesis precursor) were added. The temperature was shifted to 25°C. After growth for an additional 18 to 24 h at 25°C, the cells from 10 mL of culture were pelleted, resuspended in 1 mL methanol, sonicated for 30 min, and extracted with 4 mL hexane. After solvent evaporation, the samples were resuspended in 0.1 mL hexane and analyzed by GC-MS. In contrast to the vector-only control, _E. coli_ cells transformed with the orf880-bearing vector produced the α-olefins: 1-pentadecene and heptadecadiene (see FIG. 12). This result indicates that expression of orf880 confers on _E. coli_ the ability to biosynthesize α-olefins when grown on glucose and that the direct precursors are the most abundant fatty acids in _E. coli_, namely hexadecanoic acid and vaccenic acid (11-cis-octadecenoic acid).

Example 7

Production of α-Olefins in _E. coli_ by Heterologous Expression of *Corynebacterium efficiens* YS-134 orfCE2459

[0286] Genomic DNA encoding _C. efficiens_ YS-134 (DSMZ44549) orfCE2459 (SEQ ID NO:6) was cloned into the NcoI and EcoRI sites of vector OP80 (pCL1920 derivative) under the control of the _P_{lac}^n_ promoter. The resulting construct was then transformed into _E. coli_ C41 ΔfadE.

[0287] The _E. coli_ C41 ΔfadE was constructed using primers YafV_NorI and Ivyl_01 to amplify about 830 bp upstream of fadE and primers Lpcaf_C1 and LpR_Bam to amplify about 960 bp downstream of fadE. Overlap PCR was used to create a construct for in-frame deletion of the complete fadE gene. The fadE deletion construct was cloned into the temperature-sensitive plasmid pKOV3, which contained a sacB gene for counterselection, and a chromosomal deletion of fadE was made according to the method of Link et al., *J. Bact.* 179:6228-6237, 1997. This strain has the acyl-CoA dehydrogenase gene (fadE) deleted, therefore it cannot degrade activated fatty acids.

[0288] The cells were grown at 37°C in M9 minimal medium supplemented with 20 μg/mL uracil, 1% glucose (w/v), and 100 μg/mL spectinomycin. When the culture reached an OD_{600} of 0.8 to 1.0, it was induced with 1 mM IPTG supplemented with 0.5 mM delta-aminolaevulinic acid (heme biosynthesis precursor), and 0.05% stearic acid (as a substrate for orfCE2459). The temperature was shifted to 25°C and cells were allowed to grow for an additional 18-24 h. Cells from 1 mL of culture were resuspended with 100 μL of methanol, sonicated for 30 min, and extracted with 300 μL of ethyl acetate. After vortexing for 15 min, 300 μL of water was added, and the mixture was vortexed for another 15 min before centrifugation at 15,000 rpm for 10 min. The solvent layer was analyzed by GC-MS.

[0289] As shown in FIG. 13, _E. coli_ cells transformed with an orfCE2459-bearing vector produced 1-heptadecadiene without stearic acid feeding and produced 1-heptadecene with stearic acid feeding. This result indicates that expression of orfCE2459 confers on _E. coli_ the ability to biosynthesize α-olefins (e.g., 1-heptadecadiene) when grown on glucose (FIG. 13B) and that orfCE2459 can convert fed fatty acids, such as stearic acid, into α-olefins, such as 1-heptadecene (FIG. 13C).

Example 8

Production of α-Olefins in _E. coli_ by Heterologous Expression of *Kokuria rhizophila* orfKR121570

[0290] Genomic DNA encoding *Kokuria rhizophila* DC2201 orfKR121570 (SEQ ID NO:8), which was codon-optimized (SEQ ID NO:9) for expression in _E. coli_, was cloned into the Ndel and Xhol sites of vector OP183 (pACYC177 derivative) under the control of the _P_{lac}^n_ promoter. The resulting construct was then transformed into _E. coli_ C41 ΔfadE. The cells were grown, extracted, and analyzed as described in Example 7.

[0291] As shown in FIG. 14, _E. coli_ cells transformed with codon-optimized orfKR121570-bearing vector produced 1-heptadecadiene without stearic acid feeding and produced 1-heptadecene with stearic acid feeding. This result indicates that expression of orfKR121570 confers on _E. coli_ the ability to biosynthesize α-olefins (e.g., 1-heptadecadiene) when grown on glucose (FIG. 14C) and that orfKR121570 can convert fed fatty acids, such as stearic acid, into α-olefins, such as 1-heptadecene (FIG. 14D).

Example 9

Production of α-Olefins in *Bacillus subtilis* by Heterologous Expression of *Jeotgalicoccus* sp. ATCC8456 orf880

[0292] Genomic DNA encoding *Jeotgalicoccus* ATCC 8456_orf880 was cloned into the BamHI and XbaI sites of vector pH101 under the control of the _P_{lac}^n_ promoter. _B. subtilis_ IIIA01 (lacA::spec leuB met85 r(+)-m(+) Sp-, obtained from Bacillus Genetic Stock Center, Columbus, Ohio, strain number BGSC 1A785) was transformed with the resulting vector. The _B. subtilis_ cells were grown at 37°C in minimal medium supplemented with 2% glucose, 0.5 mM leucine, 0.3 mM methionine, and 10 mg/L chloramphenicol. When the cultures reached an OD_{600} of 0.8 to 1.0, 1 mM IPTG was added (to induce expression of orf880) along with 0.5 mM 5-amino laevulinic acid (as a precursor for heme biosynthesis) and 0.05% stearic acid (as a substrate for ORF880). The temperature was shifted to 25°C, and after growth for additional 18-24 h at 25°C, cells from a 1 mL culture were pelleted, resuspended in 100 μL methanol, sonicated for 1 h, and extracted with 300 μL ethyl acetate. 300 μL of water was added before separating the organic layer by centrifugation at 15,000 rpm for 5 min. The solvent layer was analyzed by GC-MS.

[0293] In contrast to cells transformed with the vector only control, _B. subtilis_ cells transformed with *Jeotgalicoccus* sp. ATCC8456_orf880-bearing vector produced 1-heptadecene from stearic acid (see FIG. 15B). This example demonstrates that _B. subtilis_ fatty acids, such as stearic acid, can be converted to α-olefins by heterologous expression of _Jeotgalicoccus* ATCC 8456_orf880 in _Bacillus subtilis_.
Example 10

Production of \( \alpha \)-Olefin in Bacillus subtilis by Heterologous Expression of Corynebacterium efficiens YS-134 orfCE2459

Genomic DNA encoding Corynebacterium efficiens YS-134 orfCE2459 was cloned into the BamHI and XbaI sites of vector pIT101 under the control of the P\(_{\text{lac}}\) promoter and transformed into B. subtilis 16801 cells. The resulting cells were grown, extracted, and analyzed as described in Example 9.

In contrast to cells transformed with the vector only control, B. subtilis cells transformed with Corynebacterium efficiens YS-134 orf_CE2459-bearing vector produced 1-heptadecene from stearic acid (see Fig. 15). This demonstrates that in B. subtilis fatty acids, such as stearic acid, can be converted to \( \alpha \)-olefins by heterologous expression of C. efficiens orf_CE2459.

Example 11

Production of \( \alpha \)-Olefin in Saccharomyces cerevisiae by Heterologous Expression of Jeotgalicoccus sp. ATCC8456 orf880

Genomic DNA encoding Jeotgalicoccus sp. ATCC 8456_orf880 is cloned into the Apal and Xhol sites of the yeast expression vector pESC-His. S. cerevisiae (BY4741) yeast cells are then transformed with either a plasmid containing orf880 or a plasmid not containing orf880. The transformed yeast cells are cultured under suitable conditions to allow expression of ORF880. The cells are pelleted and then lysed using YeastBuster (Novagen, Madison, Wis.). The \( \alpha \)-olefins are extracted and analyzed as described in Examples 6 and 7.

Jeotgalicoccus sp. ATCC8456_orf880 expressed in S. cerevisiae produces \( \alpha \)-olefins. Similar methods are used to express Jeotgalicoccus sp. ATCC8456_orf880 and any cell of interest, after which cell lysates are prepared and the production of \( \alpha \)-olefins analyzed.

Example 12

In Vitro Synthesis of \( \alpha \)-Olefin by Jeotgalicoccus sp. ATCC8456_orf880 Heterologously Expressed in E. coli

Expression and Purification of Jeotgalicoccus sp. ATCC8456_orf880

The genomic DNA encoding Jeotgalicoccus sp. ATCC8456_orf880 was cloned into the Ndel and Xhol sites of vector pET15b (Novagen) under the control of the T7 promoter for expression in and purification from E. coli. This plasmid expressed an N-terminal His-tagged version of 8456_orf880.

An E. coli BL21 strain (DE3) (Invitrogen) was transformed with pET15b-orf880 using routine chemical transformation techniques. Protein expression was carried out by first inoculating a colony of the E. coli strain in 5 mL of LB media supplemented with 100 mg/L of carbenicillin and shaken overnight at 37°C to produce a starter culture. This starter culture was used to inoculate 1 L of LB media supplemented with 100 mg/L of carbenicillin. The culture was shaken at 37°C until an OD\(_{600}\) value of 0.6 was reached. The culture was then placed on ice for 10 min before IPTG was added to a final concentration of 250 \( \mu \)M. The culture was then shaken at 18°C for approximately 18 h. The culture was then centrifuged at 3,700 rpm for 20 min at 4°C. The pellet was then resuspended in 30 mL of buffer containing 100 mM sodium phosphate buffer at pH 7.2 supplemented with Bacterial Protease Arrest (GBiosciences). The cells were then sonicated at 12 W on ice for 10 min. The media was then centrifuged at 3,700 rpm for 20 min at 4°C. The supernatant and the mixture was gently stirred at 4°C. The slurry was passed over a column removing the resin from the lysate. The resin was then washed with 30 mL of buffer containing 100 mM sodium phosphate buffer at pH 7.2 plus 30 mM imidazole. Finally, the protein was eluted with 15 mL of 100 mM sodium phosphate buffer at pH 7.2 plus 250 mM imidazole. The protein solution was dialyzed with 200 volumes of 100 mM sodium phosphate buffer at pH 7.2. Protein concentration was determined using the Bradford assay (BioRad). 125 \( \mu \)g/mL of protein was obtained.

In Vitro Synthesis of \( \alpha \)-Olefin from Fatty Acid Substrates

In order to assay the in vitro fatty acid substrate specificity of ORF880, potassium salts of the following fatty acids were prepared: tetradecanoic acid, hexadecanoic acid, octadecanoic acid, eicosanoic acid, and behenic acid (Sigma). The fatty acid solutions were made with 2% ethanol and 2% Tertigol solution (Sigma, St. Louis, Mo.) to a final concentration of 20 mM.

The kinetics of the decarboxylation reaction was determined using potassium octadecanoate. A 200 \( \mu \)L reaction was prepared with the following reactant concentrations: 1.25 \( \mu \)M of ORF880, 200 \( \mu \)M of potassium octadecanoate, 200 \( \mu \)L dithiothreitol, and 100 mM sodium phosphate buffer at pH 7.2. The reactions were incubated at room temperature and time points were taken in duplicate between 5 min and 120 min. Reactions were quenched and extracted by adding 100 \( \mu \)L of ethyl acetate containing 1-octadecene at 5 mg/L as an internal reference. Samples were analyzed using GC/MS using the alkane-1 splitless method, which is performed using the following parameters: run time: 20 min; column: HP-5-MS Part No. 19091S-435E (length of 30 meters; i.D.: 0.25 mm narrowbore; film: 0.25 \( \mu \)M); sample: standard ethyl acetate extraction; inject: 1 \( \mu \)L; Agilent 6850 inlet; inlet: 300°C; splitless; carrier gas: helium; flow: 1.3 mL/min; oven temp: 100°C; hold 5 min, 320 at 20°C/min, 320 hold 5 min; det: Agilent 5975B VL MSD; det. temp: 300°C; scan: 50–500 MZ. Calibration curves were generated using 1-heptadecene dissolved in ethyl acetate. Based upon this analysis, the product production was determined to be linear from 5 min to 60 min.

To assay the reaction rates of different fatty acid substrates, the following reactions were prepared at 200 \( \mu \)L each: 1.0 \( \mu \)M ORF880 enzyme, 200 \( \mu \)M of a test fatty acid salt, 200 \( \mu \)L dithiothreitol, and 100 mM sodium phosphate buffer at pH 7.2. The reactions were carried out at room temperature and time points were taken in triplicate at 20 min and 45 min using the extraction and analysis procedure described above. Reference curves were generated using available chemical standards. In some instances, the chemical standards were not available (for example, cis-9-heneicosene was used as a reference for 1-heneicosene and 9-tricosene as used as a reference for 1-tricosene). Activi-
ties were calculated by taking the difference between the average α-olefin concentrations for each substrate at 47 min and 20 min and then dividing the difference by 27 min. The results are summarized in Table 7.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Activity (nM alkene produced/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tetradecanoic</td>
<td>22.9</td>
</tr>
<tr>
<td>hexadecanoic</td>
<td>181.9</td>
</tr>
<tr>
<td>octadecanoic</td>
<td>77.2</td>
</tr>
<tr>
<td>eicosanoic acid</td>
<td>19.7</td>
</tr>
<tr>
<td>behenic acid</td>
<td>30.6</td>
</tr>
</tbody>
</table>

These results demonstrate that heterologously expressed ORF880 was able to convert various fatty acid substrates to α-olefins in vitro. The data also show that ORF880 had highest activity when hexadecanoic acid was the fatty acid substrate.

Example 13

In Vitro Synthesis of α-Olefins by Corynebacterium efficiens YS-134 orfCE2459 Heterologously Expressed in E. coli

An E. coli BL21 strain (DE3) (Invitrogen) was transformed with pET15-orf CE2459 using routine transformation techniques. Protein expression was carried out by first inoculating a colony of the E. coli strain in 5 mL of LB media supplemented with 100 mg/L of carbencillin and then shaking overnight at 37°C, to produce a starter culture. This starter culture was used to inoculate 0.5 L of LB media supplemented with 100 mg/L of carbencillin. The culture was shaken at 37°C until an OD600 value of 0.6 was reached. The culture was placed on ice for 10 min before IPTG was added to a final concentration of 250 μM. δ-Aminolevulinic acid was also added to a final concentration of 0.5 mM. The culture was then shaken at 25°C for approximately 18 h. The culture was then centrifuged at 3,700 rpm for 20 min at 4°C. The pellet was then resuspended in 7 mL of buffer containing 100 mM sodium phosphate buffer at pH 7.2 supplemented with Bacterial Protease Arrest (GBiosciences). The cells were then sonicated at 12 W on ice for 9 sec with 1.5 sec of sonication followed by 1.5 sec of rest. This procedure was repeated 4 times with one min intervals between each sonication cycle. The cell-free extract was centrifuged at 10,000 rpm for 30 min at 4°C.

To assay the activity of ORF CE2459, 200 μL of lysate was incubated with 200 μM of hydrogen peroxide and 200 μM of potassium octadecanoate. As a negative control, the same reaction was set up, but without the addition of either cell lysate (100 mM sodium phosphate pH 7.2 was used in its place) or free fatty acid. The reactions were incubated at room temperature overnight. The reactions were then extracted with 100 μL of ethyl acetate with 1-olectadecene added at 5 mg/L as an internal standard. Trimethylsilylamine hydroxide was added to the ethyl acetate layer at a 1:1 ratio and the extracts analyzed by GC/MS using the MAR_splittless_short method. The parameters used were as follows: 9.50 min; column: DB-35, 0.18 x 30 m, 0.25 μm; sample: standard ethyl acetate extraction; inject: 1 μL. Agilent 6850 inlet, inlet: 300°C. splitless, carrier gas: helium; flow: 1.3 mL/min; oven temp: 100°C, hold 5 min, 320 at 20°C/min, 320 hold 5 min; det: Agilent 5975B VL MSD; det. temp: 300°C; scan: 50-500 M/Ze.

As shown in FIG. 17, an E. coli lysate expressing orfCE2459 was capable of converting octadecanoic acid into 1-olectadecene. The peak at 5.20 min corresponds to 1-olectadecene and the peak at 8.76 is the internal standard 1-olectadecene. Peak identification was determined by comparing the retention time and the MS fragmentation pattern with authentic references from Sigma. This result demonstrates that ORF CE2459 exhibits similar in vitro fatty acid decarboxylase activity as ORF880.

**Example 14**

In Vitro Synthesis of α-Olefins by Kokuria rhizophila orf_KRH21570 Heterologously Expressed in and Purified from E. coli

The genomic DNA encoding Kokuria rhizophila orf_KRH21570 was cloned into the NdeI and XhoI sites of vector pET15b (Novagen) under the control of the T7 promoter for expression in and purification from E. coli. This plasmid expressed an N-terminal His-tagged protein.

E. coli C41 DE3 (A1adE) cells were transformed with pET15b-orfK21570 using routine transformation techniques. Protein expression was carried out by first inoculating 5 mL of LB broth supplemented with 100 mg/L carbencillin with a colony of the E. coli strain which was followed by shaking the cells overnight at 37°C to produce a starter culture. This starter culture was used to inoculate 0.5 L of LB media supplemented with 100 mg/L of carbencillin. The culture was shaken at 37°C until an OD600 value of 0.6 was reached, after which it was induced with 1 mM IPTG and 0.5 mM delta-aminolevulinic acid. The culture was then shaken at 25°C for approximately 18 h. The protein was purified as described in Example 12.

To assay the activity of ORF_KRH21570, 200 μL of purified protein was incubated with 200 μM of hydrogen peroxide and 200 μM of potassium hexadecanoate. As a negative control, the same reaction was conducted, but without the addition of either purified protein (100 mM sodium phosphate pH 7.2 was used in its place) or free fatty acid. The reactions were performed in duplicates and incubated at room temperature overnight. The reactions were then extracted with 100 μL of ethyl acetate with 1-olectadecene added at 5 mg/L as an internal standard. Trimethylsilylamine hydroxide was added to the ethyl acetate layer at a 1:1 ratio and the extracts analyzed by GC/MS using the MAR_splittless_short method. The parameters used were as follows: 9.50 min; column: DB-35, 0.18 x 30 m, 0.25 μm; sample: standard ethyl acetate extraction; inject: 1 μL. Agilent 6850 inlet, inlet: 300°C. splitless, carrier gas: helium; flow: 1.3 mL/min; oven temp: 80°C, hold 1.0 min, 320 at 30°C/min, 320 hold 0.5 min; det: Agilent 5975B VL MSD; E1 mode; det. temp: 230°C; scan: 50-330 M/Ze.

As shown in FIG. 18, Kokuria rhizophila ORF_KRH21570 was capable of converting hexadecanoic acid to 1-olectadecene in vitro.
Example 15

In Vitro Synthesis of α-Olefins by 
*Methylobacterium populi* orf_Mpop1292 
Heterologously Expressed in and Purified from *E.
coli* 

[0311] The genomic DNA encoding *Methylobacterium populi* orf_Mpop1292 (SEQ ID NO:11) was cloned into the NdeI and XhoI sites of vector pET15b (Novagen) under the control of the T7 promoter for expression in and purification from *E. coli*. This plasmid expressed an N-terminal Histagged version of orf_KR121570.

[0312] *E. coli* C41 DE3 (A(fadE)) cells were transformed with pET15b-orf_Mpop1292 using routine transformation techniques. Protein expression was carried out by first inoculating 5 mL of LB broth supplemented with 100 mg/L carbenicillin with a colony of the *E. coli* strain, and then shaking the cells overnight at 37°C to produce a starter culture. This starter culture was used to inoculate 0.5 L of LB media supplemented with 100 mg/L of carbenicillin. The culture was shaken at 37°C until an OD_{600} value of 0.6 was reached, after which it was induced with 1 mM IPTG and 0.5 mM delta-aminolaevulinic acid. The culture was then shaken at 25°C for about 18 h. The protein was purified as described in Example 12 and assayed as described in Example 14.

[0313] As shown in FIG. 19, *Methylobacterium populi* ORF_Mpop1292 was capable of converting hexadecanoic acid to 1-pentadecanone in vitro.

Example 16

Production of α-Olefins in *E. coli* Through 
Heterologous Expression of *Bacillus subtilis* Fatty 
Acid Hydrolase, P450<sub>orfK</sub> 

[0314] The genomic DNA encoding *Bacillus subtilis* fatty acid hydrolase, Cyp152A1 (SEQ ID NO:13) or P450<sub>orfK</sub>, was cloned into the NdeI and XhoI sites of vector OP-183 (pACYC177 derivative) under the control of the P<sub>vas</sub> promoter. The vector was used as a template to introduce a mutation at position 85 changing the Gln to His using the following primers: GTTAATGCGATACCGAATGGATGGGTC. The resulting constructs were then transformed into *E. coli* C41 (A(fadE)). The cells were grown at 37°C in M9 minimal media supplemented with 20 μg/mL uracil, 1% glucose (w/v), and 100 μg/mL carbenicillin. When the culture reached OD<sub>600</sub> of 0.8-1.0, it was induced with 1 mM IPTG; supplemented with 0.5 mM delta-aminolaevulinic acid (heme biosynthesis precursor) and 0.05% stearic acid (as a substrate for B<sub>sh</sub>). The temperature was shifted down to 25°C and cells were allowed to grow for an additional 18-20 h. Cells from 1 mL of culture was resuspended with 100 μL of methanol, sonicated for 30 min, and extracted with 300 μL of ethyl acetate. After vortexing the extract for 15 min, 300 μL of water was added, vortexed for another 15 minutes before centrifugation at 15,000 rpm for 10 minutes. The solvent layer was analyzed by GC-MS.

[0315] As shown in FIG. 20, *E. coli* cells transformed with *Bacillus subtilis* P450<sub>orfK</sub>-bearing vectors produced 1-heptadecene with stearic acid feeding. This result indicates that *Bacillus subtilis* P450<sub>orfK</sub> can convert feed fatty acids, such as stearic acid, into α-olefins, such as 1-heptadecene (FIG. 20).

Both the wildtype protein and the Gln85-His-mutant protein were capable of carrying out this reaction.

Example 17

Specific Activity of α-Olein Production and α and 
β Hydroxy Fatty Acid Production Using ORF880, 
CYP152A1 (P450<sub>orf</sub>), and CYP152A1 Q85H 

[0316] The specific activity of α-olein formation as well as α- and β-(3)hydroxy fatty acid formation were determined for ORF880, CYP152A1, and CYP152A1 Q85H (see Examples 12 and 16). An *E. coli* BL21 strain (DE3) (Invitrogen) was transformed with either pET15b-orf 880, pET15b-cyp152A1, and pET15b-cyp152A1 Q85H using routine chemical transformation techniques. Protein expression was carried out by first inoculating a colony of the *E. coli* strain in 5 mL of LB media supplemented with 100 mg/L of carbenicillin and shaken overnight at 37°C to produce a starter culture. This starter culture was used to inoculate 0.5 L of LB media supplemented with 100 mg/L of carbenicillin. The culture was shaken at 37°C until an OD<sub>600</sub> value of 0.6 was reached. The culture was placed on ice for 10 min before IPTG was added to a final concentration of 250 μM. The culture was shaken at 25°C for approximately 18 h, and then centrifuged at 3,700 rpm for 20 min at 4°C. The pellet was then resuspended in 10 mL of buffer containing 100 mM sodium phosphate buffer at pH 7.2 supplemented with Bacterial Proteinase/Anti G-Biosciences. The cells were then sonicated at 12 W on ice for 9 sec with 1.5 sec of sonication followed by 1.5 sec of rest. This procedure was repeated 5 times with one min intervals between each sonication cycle. The cell free extract was centrifuged at 10,000 rpm for 30 min at 4°C. 5 mL of Ni-NTA (Qiagen) was added to the supernatant, and the mixture was gently stirred at 4°C. The slurry was passed over a column removing the resin from the lysate. The resin was then washed with 30 mL of buffer containing 100 mM sodium phosphate buffer at pH 7.2 plus 30 mM imidazole. Finally, the protein was eluted with 15 mL of 100 mM sodium phosphate buffer at pH 7.2 plus 250 mM imidazole. The protein solution was dialyzed with 200 volumes of 100 mM sodium phosphate buffer at pH 7.2 with 20% glycerol. Protein concentration was determined using the Bradford assay (Biorad). 578 μg/mL of ORF880, 865 μg/mL of CYP152A1, and 653 μg/mL of CYP152A1 Q85H protein was obtained.

[0317] To determine the specific activity with potassium hexa decanolate, the following reactions were set up where each time point consisted of 500 μL. Each reaction contained of 500 μM K<sub>2</sub> <sub>3</sub> <sub>H</sub> <sub>2</sub> <sub>O</sub>, 200 μM of potassium hexadecanolate, 100 mM sodium phosphate buffer at pH 7.2, and protein at the following concentrations: ORF880-0.23 or 0.46 μM; CYP152A1-8.65, 17.3, or 34.6 μM; CYP152A1 Q85H-6.53, 13.06, or 26.12 μM. The reactions were carried out at room temperature and time points were taken at 0, 1, 3, 5, 7, 10, 14, and 20 minutes. Reactions were quenched with 50 μL of 12 M HCl and then extracted with 200 μL of ethyl acetate. 50 μL of this extract was then reacted with 50 μL N, O-Bis(trimethylsilyl) trifluoroacetamide with trimethylchlorosilane (Aldrich). Samples were analyzed using GC/MS using the following parameters: run time: 20 min; column: HP-S-MS Part No. 19091S-433E (length of 30 meters; i.d.: 0.25 mm narrowbore; film: 0.25 μM); inject: 1 μL Agilent 6850 inlet; inlet: 300°C splitless; carrier gas: helium; flow: 1.3 mL/min; oven temp: 100°C hold 5 min; 320°C at 20°C
C./min, 320° C. hold 5 min; det: Agilent 5975B VL MSD; det. temp: 300° C.; scan: 50-500 M/Z. Calibration curves were generated using 1-pentadecene, β-(2)-hydroxyhexadecanoic acid, and β-(3)-hydroxyhexadecanoic acid dissolved in ethyl acetate and derivatized as described herein.

**[0318]** Based upon this analysis, product formation was determined to be linear within the first three minutes. The specific activities for each enzyme concentration in the linear range were averaged for a given reaction (e.g., decarboxylation, α-hydroxylation, or β-hydroxylation). FIG. 21 summarizes the results. The specific activity for 1-pentadecene formation was similar for ORF880 and CYP152A1. Mutating the glutamine 85 residue to a histidine increased the specific activity of 1-pentadecene formation (i.e., increased fatty acid decarboxylase activity). This mutation also increased the rate of β-hydroxyhexadecanoic acid formation and lowered the rate of α-hydroxyhexadecanoic acid formation as compared with the wild-type version of CYP152A1. The rate of hydroxyhexadecanoic acid production by ORF880 was about 1/4th of the rates exhibited by the two CYP152A1 enzymes. The ratio of 1-pentadecene formation (i.e. decarboxylation) to hydroxyhexadecanoic acid formation (i.e., hydroxylation) activities for each enzyme is given in FIG. 22.

**Example 18**

Analysis of ORF880 Homolog Activity

**[0319]** As described in Examples 7 through 17, five P450 enzymes were able to convert fatty acids into α-olefins in-vivo and in-vitro, and three of these P450 enzymes were also able to convert fatty acids into β-hydroxy fatty acids (see Table 8). In contrast, two P450 enzymes, *Bacillus clausii* orf ABC3040 and *Sphingomonas paucimobilis* P450<sub>300</sub>, were not able to make α-olefins or β-hydroxy fatty acids. These P450 enzymes converted fatty acids only to the corresponding α-hydroxy fatty acids (see Table 8). Consequently, there appears to be a correlation between the ability of a P450 enzyme to hydroxylate fatty acids in the β-position and to decarboxylate fatty acids to the α-olefin, and the inability of a P450 enzyme to decarboxylate fatty acids to α-olefins if the enzyme can only hydroxylate fatty acids in the β-position. In other words, P450 enzymes that hydroxylate fatty acids in the β-position also have the potential to decarboxylate these fatty acids to α-olefins.

**[0320]** Table 8 also demonstrates that when a His or Met is at position 85, the enzyme has the ability to produce α-olefins. Table 8 also shows the percent identity of each enzyme tested relative to *Jostagallicoccus* sp. 8456 ORF880.

<table>
<thead>
<tr>
<th>ORF880 Homolog Tested</th>
<th>Position 85&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Percent Identity</th>
<th>Makes α-olefin&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Makes α-hyFA&lt;sup&gt;4&lt;/sup&gt;</th>
<th>Makes β-hyFA&lt;sup&gt;4&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Jostagallicoccus</em> sp. 8456 orf880</td>
<td>His</td>
<td>100</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><em>Corynebacterium efficiens</em> NP_739069&lt;sup&gt;1&lt;/sup&gt;</td>
<td>His</td>
<td>27</td>
<td>+</td>
<td>(-)</td>
<td>(-)</td>
</tr>
<tr>
<td><em>Kokualia thiophilum</em> YP_001856010&lt;sup&gt;1&lt;/sup&gt;</td>
<td>His</td>
<td>29</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><em>K. thiophilum</em> orf_KR2121570</td>
<td>Gln</td>
<td>37</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><em>Bacillus clausii</em> YP_176535&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Met</td>
<td>31</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Methylotacterium populi</em></td>
<td>Gln</td>
<td>41</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Sphingomonas paucimobilis</em> P450&lt;sub&gt;300&lt;/sub&gt;</td>
<td>Gln</td>
<td>36</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>1</sup>in-vitro experiments with lysate, not with purified protein, difficult to detect hydroxyl fatty acids

<sup>2</sup>poorly expressed in *E. coli*, activity was only detected in-vitro

<sup>3</sup>Relative to the position in ORF880

<sup>4</sup>with hexadecanoic acid as substrate (in-vitro and in-vivo)

**hyFA** = hydroxy fatty acid

**OTHER EMBODIMENTS**

**[0321]** It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

**SEQUENCE LISTING**

<160> NUMBER OF SEQ ID NOS: 14
<210> SEQ ID NO 1
<211> LENGTH: 1269
-continued

<210> SEQ ID NO 2
<211> LENGTH: 422
<212> TYPE: PRT
<213> ORGANISM: Jeotgalicoccus sp. ATCC8456

<400> SEQUENCE: 2

Met Ala Thr Leu Lys Arg Asp Gly Leu Asp Asn Thr Leu Lys Val
1  5  10  15
Leu Lys Gln Gly Tyr Leu Tyr Thr Thr Asn Glu Arg Asn Arg Leu Asn
20  25  30
Thr Ser Val Phe Gln Thr Lys Ala Leu Gln Gly Lys Pro Phe Val Val
35  40  45
Val Thr Gly Lys Glu Gly Ala Glu Met Phe Tyr Asn Asp Val Glu Val
50  55  60
Gln Arg Glu Gly Met Leu Arg Lys Ile Val Asn Thr Leu Phe Gly
65  70  75  80
Lys Gly Ala Ile His Thr Val Asp Gly Lys Lys His Val Arg Arg Lys
85  90  95
Ala Leu Phe Met Ser Leu Met Thr Glu Gly Arg Asn Arg Tyr Val Arg
Glu Leu Thr Arg Thr Leu Trp His Ala Asn Thr Gln Arg Met Glu Ser
110
115
120
125
Met Asp Glu Val Asn Ile Tyr Arg Glu Ser Ile Val Leu Leu Thr Lys
130
135
140
Val Gly Thr Arg Trp Ala Gly Val Gin Ala Pro Pro Glu Asp Ile Glu
145
150
155
160
Arg Ile Ala Thr Asp Met Asp Ile Met Ile Asp Ser Phe Arg Ala Leu
165
170
175
Gly Gly Ala Phe Lys Gly Tyr Lys Ala Ser Lys Glu Ala Arg Arg Arg
180
185
190
Val Glu Asp Trp Leu Glu Glu Gin Ile Ile Glu Thr Arg Lys Gly Asn
195
200
205
Ile His Pro Pro Glu Gly Thr Ala Tyr Glu Phe Ala His Trp Glu
210
215
220
Asp Tyr Leu Gly Asn Pro Met Asp Ser Arg Thr Cys Ala Ile Asp Leu
225
230
235
240
Met Asn Thr Phe Arg Pro Leu Ile Ala Ile Asn Arg Phe Val Ser Phe
245
250
255
Gly Leu His Ala Met Asn Glu Asn Pro Ile Thr Arg Glu Lys Ile Lys
260
265
270
Ser Glu Pro Asp Tyr Ala Tyr Lys Phe Ala Gin Glu Val Arg Arg Tyr
275
280
285
Tyr Pro Phe Val Pro Phe Leu Pro Gly Lys Ala Lys Val Asp Ile Asp
290
295
300
Phe Gin Gly Val Thr Ile Pro Ala Gly Val Gin Leu Ala Leu Asp Val
305
310
315
320
Tyr Gly Thr Thr His Asp Glu Ser Leu Trp Asp Asp Pro Asn Glu Phe
325
330
335
Arg Pro Glu Arg Phe Glu Thr Trp Asp Gly Ser Pro Phe Asp Leu Ile
340
345
350
Pro Gin Gly Gly Gly Asp Tyr Trp Thr Asn His Arg Cys Ala Gly Glu
355
360
365
Trp Ile Thr Val Ile Ile Met Glu Thr Met Lys Tyr Phe Ala Glu
370
375
380
Lys Ile Thr Tyr Asp Val Pro Glu Gin Asp Leu Glu Val Asp Leu Asn
385
390
395
400
Ser Ile Pro Gly Tyr Val Lys Ser Gly Phe Val Ile Lys Asn Val Arg
405
410
415
Glu Val Val Asp Arg Thr
420

<210> SEQ ID NO 3
<211> LENGTH: 1507
<212> TYPE: DNA
<213> ORGANISM: Jeotgalicoccus sp. ATCC8456
<400> SEQUENCE: 3

gggtacctag ttacgacct ctaccacact acctttggcc cttttccatt
60
aagtttagtc cacgggtacct aggttgttayc gactttgcgt gttgacgggc cgggtggtgac
120
aagaccgagac gctacagcct ccggtacgtg ctgatctagc aattctagcg attcagctt
180
-continued

catggagtcc agtgtgcagac tccaatcnga actgaagaca gttttatgga atctgccttg 240
cctcggggt ctcgcactgt tggtaactctg ccaattgagt caggtgtgtc gccaaatga 300
taaggggtcat gatgtgttga cgtcacccca cacttctctc gttttgtcag cgccgagctaa 360
tctagagtcc caaagctgta gatggcaacat aaatgtaagg gttgccccct tgtggtgcac 420
taaccacaca ttcgcaagctg aagacatgag acctaacgagct ctctgcccaca 480
aaggggaaac ccaattctctt tgggacatag aggtgctcaca gattttgataa gttccttcgccc 540
gttgtccgta attaaccac atgtcctccac gttggtgctgg gtccccccca ataacctttga 600
gttaacactt tcggggtcctt cccccccaaga ggaatgctttt atgggattac tggagcaactg 660
agggggcgaa ccacccccgaacet cctgctttcgt cgggtgctta ccaaaagcttt 720
cctgtcctgt tttggtccct cccagcgcct tggagcagcag tggagcagcag 780
cctgctgca cctggggccc tctcatacttc agtagccattt actgataaagccc atatcag 840
actctcccc ttcgcctcag atgaaacacgc ttcgactgta gttgcccggg 900
gggttcaaatg tctagactcag tctagcctgtg acggccttgga ccatgggagat 960
aaacgctgcg acctagactcag tccagtctggct gttgacagctg ggtgcttcgct 1020	
tttaactccg atcatctcag gcggctttcatt tccttcctca acaacagagt 1080
ttttaagacgc gaaaacctcc tcacatctctag cgggtgctgtg ccgctgcactg 1140
gggaggcttt cctattgttg gctgcccttgag ggattgctgg cgccctgcttg 1200
gcctccagt cccagtcggt ggtggactgc tctcctccac cctgggcaac 1260
acctatgtg cccggtcggc cccatctcttt cctgacagat aaatcagccct ttacttaaga 1320	ttactctcttg aatcagggctt atgggcttctt cccagcctgc ttcctgagct 1380
aaagggggtg tgtgcgggca ttgatcctcc gctgcgagcc tccgctctga ggagggagct 1440
cctgacgctg tggagtgtctg agtagcagcc ccagggctcag ttcgagccca 1500
ggagagcg 1507

<210> SEQ ID NO 4
<211> LENGTH: 1209
<212> TYPE: DNA
<213> ORGANISM: Jeotgalicoccus sp. ATCC6456
<400> SEQUENCE: 4

atggtctact tgaagaagttg caaaggtctg gataacacctg tgaagttctg gaaacaaagtt 60
taactctgca ctacacccag cacagacagct ctgaccaccgc cgtctctctca aaccaaaaggc 120
cctggggtcag aaacgcttgct ggtgtgctga ggcaaaagaa gagcgagagat gttctatatc 180
aaaggttagtgg tacagcctgtgg gggcatgcctg ccggacdgcata tgtaaacacg cttgtctgggc 240
aaggggtgctg ctatacctgtg gatggccagac aaacgtagag acgcgtataac agctgtctcattg 300	tctctgtatggtgaggagca cctgacactg tcatgtggagc tcgcctcgcag cctgtgggcat 360
gcgaacgacg agcggtatgga atctagtgtg gagggtacatca tctacgctga aagcatcttgt 420
tgtcgtacgag aaggtcgccgca cctgcgctgc gggtgtccgag ccggcggagag gacattggag 480
goatcgctgct cggacatgcc tgtatctctggt gtgctctggg ggccgcaatct 540
atcctgggaa cccgttaaag taaatcctccac cccgggagag gtcgggtctt gtagaatattt 600
gccacagggg aggattactt ggtaatacctg aaggtactgtc gtcctggcag gatcagcttg 660
atgaacagt ttgcgcgcgt gatgccttc aaccgcatgg ttctcttttcgg ttgctcgccgct

atgaacgaaa acccgatcac togtgagaag ... acctggaaagt agatctgaac

tccatc.ccgg gctatgctaa aag...ggcttt gttatcaaaa acgt.ccgtga agtag
cgcaccitaa

<210> SEQ ID NO 5
<211> LENGTH: 422
<212> TYPE: PRT
<213> ORGANISM: Jeotgalicoccus sp. ATCC8456

<400> SEQUENCE: 5

Met Ala Thr Leu Lys Arg Asp Lys Gly Leu Asp Asn Thr Leu Lys Val
1  5   10  15
Leu Lys Gln Gly Tyr Leu Tyr Thr Thr Asn Gln Arg Arg Leu Asn
20  25   30
Thr Ser Val Phe Gln Thr Lys Ala Leu Gly Gly Lys Pro Phe Val Val
35  40   45
Val Thr Gly Lys Glu Gly Ala Glu Met Phe Tyr Asn Arg Asp Val Val
50  55   60
Gln Arg Glu Gly Met Leu Pro Lys Arg Ile Val Asn Thr Leu Phe Gly
65  70   75   80
Lys Gly Ala Ile His Thr Val Asp Gly Lys Lys His Val Asp Arg Lys
85  90   95
Ala Leu Phe Met Ser Leu Met Thr Glu Gly Asn Leu Asn Tyr Val Arg
100  105  110
Glu Leu Thr Arg Thr Leu Trp His Ala Asn Thr Gin Arg Met Glu Ser
115  120  125
Met Asp Glu Val Asn Ile Tyr Arg Glu Ser Ile Val Leu Leu Thr Lys
130  135  140
Val Gly Thr Arg Trp Ala Gly Val Gin Ala Pro Pro Glu Asp Ile Glu
145  150  155  160
Arg Ile Ala Thr Asp Met Asp Ile Met Ile Asp Ser Phe Arg Ala Leu
165  170  175
Gly Gly Ala Phe Lys Gly Tyr Lys Ala Ser Lys Gly Ala Arg Arg Arg
180  185  190
Val Glu Asp Trp Leu Glu Glu Gin Ile Ile Glu Thr Arg Lys Gly Asn
195  200  205
Ile His Pro Pro Glu Gly Thr Ala Leu Tyr Glu Phe Ala His Trp Glu
210  215  220
Asp Tyr Leu Gly Asn Pro Met Asp Ser Arg Thr Cys Ala Ile Asp Leu
225  230  235  240
Met Asn Thr Phe Arg Pro Leu Ile Ala Ile Asn Arg Phe Val Ser Phe
245  250   255
| Gly Leu His Ala Met Asn Glu Asn Pro Ile Thr Arg Glu Lys Ile Lys  |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 260               | 265               | 270               |
| Ser Glu Pro Asp Tyr Ala Tyr Lys Phe Ala Gin Glu Val Arg Arg Tyr |
| 275               | 280               | 285               |
| Tyr Pro Phe Val Pro Phe Leu Pro Phe Lys Lys Val Asp Ile Asp |
| 290               | 295               | 300               |
| Phe Gin Gly Val Thr Ile Pro Ala Gly Val Gly Leu Ala Leu Asp Val |
| 305               | 310               | 315               | 320               |
| Tyr Gly Thr Thr His Asp Glu Ser Leu Thr Trp Asp Asp Pro Am Glu Phe |
| 325               | 330               | 335               |
| Arg Pro Glu Arg Phe Glu Thr Trp Asp Gly Ser Pro Phe Asp Leu Ile |
| 340               | 345               | 350               |
| Pro Gin Gly Gly Gly Asp Tyr Trp Thr Thr His Arg Cys Ala Gly Glu |
| 355               | 360               | 365               |
| Trp Ile Thr Val Ile Met Glu Thr Met Lys Thr Phe Ala Glu |
| 370               | 375               | 380               |
| Lys Ile Thr Tyr Asp Val Pro Glu Gin Asp Leu Glu Val Asp Leu Ann |
| 385               | 390               | 395               | 400               |
| Ser Ile Pro Gly Tyr Val Lys Ser Gly Phe Val Ile Lys Ann Arg |
| 405               | 410               | 415               |
| Glu Val Val Asp Arg Thr  |
| 420               |

<210> SEQ ID NO 6
<211> LENGTH: 1308
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium efficiens

<400> SEQUENCE: 6
ttcgggttcc aacgcaaaccc gcagctcgggt tctggagagcg gtagcatctc tgggccagg
| 60 |
gaaacgggta tcacgcgtcga cccctggagag cacaagccac ggcgcgcagta aagcctcaac|
| 120 |
catgggcttg agcgcgggca tggcagatctt ctcgcggggag cagcgggtgac ccggctaac|
| 180 |
cggggcggg cccctggagac cgaaggtttgt gaggctctca tagtctctct cgggggccag|
| 240 |
gtctgctccgg ggacgagggac gtctcgctgtg aacgctctgc ggctcttccc actctatggg|
| 300 |
gtctgcttgg gtgcgcttga tggctgtatgc ccacgctgca cccctcatgca cgggggccac|
| 360 |
ctgcttttgg gggagaaggg tggcagctgc cggcagctag ggcacaaag gataagcgcg|
| 420 |
ccgggtctcc tggcctgaag cgaagggcag gggcgctctc ccccttgagga tctcttccac|
| 480 |
cacgggggg gtcggaggtca ggaccgtcat gggcgaggtgc cggggccaga gtaactgcac|
| 540 |
gtgggggggt ggtcgagttgt gtaattccat gcggcctatgg ggcggctgca cggcctccc|
| 600 |
gtggggggggacaccaggg ggctgggtctc cgggctgtca cgggctttca cgggctttca|
| 660 |
ccggcgccgc gcctggcctga ttgctttcag ggccccactgg tcaccagccg cgggggtgttct|
| 720 |
ccacgagaggc gctggcttctt tcggcccttg gacgactgta tagaccgact ggcctctgctg|
| 780 |
atggggggcgc ggctggctgtct cttggctgct eacgctaatgc cggcctccac ggttagggccgc|
| 840 |
aacgcccaag gcggcagcgc cggcctgtca ggcagcgggg ggttgccggg cccagtctcg|
| 900 |
ccagacccgg tacccctgac ggcggagcag tgcagctgac ccggcactggt ccctgtgctc|
| 960 |
atagggcaca gtcggctctcg gacggtgtgc gacagcgct cgctggctgt ccagcgaatg|
| 1020 |
ccagcggaccc tcacccgacca gggggtcgg gcataccgcg gccatgggcc cgtaacgctt|
| 1080 |
cattcgggtca ttgctataga acagctccac tcggctgtaa ccgcgcacga tggcagcggg
1140
ttggaacacg atgcccagcc gcagcggggt gtggcatacg gttgagatct ccgaccctg
1200
ggcacacagg gagagaaaaa gttacccgtg gcgcacagg tggggaacct gtggcagcggg
1260
ggcacaaggg cggagagtg tctgagcttc ccgagagaccc tcttcccc
1320

<210> SEQ ID NO: 7
<211> LENGTH: 435
<212> TYPE: PRT
<213> ORGANISM: Corynebacterium efficiens
<400> SEQUENCE: 7
Met Glu Glu Val Pro Pro Met Thr Gin Thr Ser Ser Cys Pro Phe Ala
1  5  10  15
Pro Gly Glu Gin Ala Pro Asn Leu Leu Arg His Gly Tyr Leu Phe Leu
20  25  30
Ser Arg Leu Arg Arg Lys Ala Gly Ile Ser Pro Asp Ala Asn Thr Pro
35  40  45
Leu Arg Ser Arg Met Leu Phe Lys Pro Val Thr Ile Val Arg Gly Ser
50  55  60
Ala Gly Val Glu Leu Phe Tyr Asp Asn Asp Arg Met Lys Arg Asp Gly
65  70  75  80
Ala Met Pro Ala Val Ile Arg Ile Pro Leu Phe Gly Glu Gly Ala Val
85  90  95
His Ser Leu Arg Asp Gly Glu Glu His Arg Leu Arg Lys Arg Gin Leu Ala
100 105 110
Asp Val Ala Tyr Asp Asp Lys Val Ala Glu Phe Asp Ala Leu Val
115 120 125
Arg Arg Glu Val Asp Arg Val Glu Asp Trp Ala Arg Glu Pro Gly
130 135
Thr Val Tyr Asp Gly Ala Ala Leu Ala Phe Gly Arg Ala Ala Tyr Arg
145 150 155 160
Trp Ala Gly Ile Glu Leu Ser Gin Lys Glu Ala Ser Arg Arg Ala His
165 170 175
Gln Met Ala Glu Leu Val Tyr Gin Phe Gly His Pro Leu Lys Gly His
180 185 190
Ala Leu Gly Trp Ile Asn Arg Ala Arg Leu Asn Arg Trp Ala Leu Lys
195 200 205
Leu Ile Arg Gin Ala Arg Ala Gln Arg His Val Ala Pro Gly Ser
210 215 220
Ala Leu Glu Ala Met Ser Arg Leu Val Gly Pro Asp Gly Glu Leu Val
225 230 235 240
Asp Ala Ser Ile Ala Gly Ile Glu Leu Gin Arg Leu Thr Arg Leu Pro Thr
245 250 255
Val Ala Val Ser Leu Phe Ala Ser Phe Ala Gly Ser Ala Leu Val Glu
260 265 270
His Pro Glu Trp Val Glu Lys Ile Arg Glu Gly Glu Gin Pro Val Ala
275 280 285
Phe Ala Phe Ala Gin Glu Val Arg Val Tyr Pro Phe Val Pro Met
290 295 300
Leu Pro Ala Ile Ala Thr Thr Thr Glu Ile Gin Gly Cys Pro Val
305 310 315 320
-continued

His Glu Gly Glu Arg Val Ile Ile Asp Ile Tyr Glu Thr Asn Thr Asp 326 330 335
Pro Asn Glu Trp Glu Asn Pro Ser Ala Phe Gin Pro Glu Arg Phe Leu 340 345 350
Ser Arg Glu Asp Leu Gly Thr Gin Glu Arg Tyr Glu Arg Leu Thr Ser 355 360 365
Phe Val Pro Gin Gly Gly Ala Gly Val Tyr Thr Gly His Arg Cys Pro 370 375 380
Gly Glu Lys Ile Ala Met Ala Ala Leu Thr Ala Met Val Glu Ala Leu 385 390 395 400
Cys Arg Pro Gly Val Val Leu Ser Thr Asp Pro Ala Asp Thr Arg Phe 405 410 415
Pro Trp Thr Gin Met Leu Thr Arg Ser Glu Thr Gly Met Arg Val Arg 420 425 430
Val Glu Arg 435

<210> SEQ ID NO. 8
<211> LENGTH: 1297
<212> TYPE: DNA
<213> ORGANISM: Kokuria rhizophila

<400> SEQUENCE: 8

atgacctcac cgctcggtca gaacgcgtcc gacggaggcc cgtccctacct cgctccgggc 60
tacctttag ccctcgccgc acagcgcgc gcgggctctt ctcctgcaact ggggtgcoccc 120
gtcctgcagct ctcgcttgag ccagcgcacc gtctctggtcc gcggcgagga gggcgctcaag 180
cctcttacag acacccctcgc cgtccgggcc gcagccgcca tcggcgagat cgtgacgggg 240
cgctcttccc gcgctggggc cggcccgccc tgtggcgcgg gacgcccaag 300
acaccaactg ccgagcatgc ctgagccgac gacgcctggcc cgccccccctt cacaccctgc 360
ccgagggcc tcgagaggcc cggtgcggcc tggagaggg gcgaataacat ctaagacgc 420
acccgcatcg cctccgcggcc cgctccctct ctcctgctgag gcggcgctcc ggccgagg 480
gagatggcc cgctggccgc cccctcggcc ctgctggcc gcctctctcg gcccggccc 540
acaccaacctg ttcctcgcact ggcaccgcatc ctcctgctgc ggcggggaga cggccggttc 600
agagacgcgc gcggggccctt gtcacagcg cacggagagg gcgccaggac gcctctgcatcgc 660
gccctggttg acagacgccg cgaagtcggt gcggcggaga cggggccatt cacagtcgc 720
aacctccgcc ccgaggaacct ggccgtccgg cctgctggcc ctcctgccct cccgctccctg 780
ctggagacgc ccgagcttcgc ggcagcgcct cggccgctct cggagcagc ccggccgaccc 840
cgtgctggc tcggccggtc cggagcgcct cggccgctct cggagcagc ccggccgaccc 900
ctggccatgc tcggccggtc ggctccagac gcacggcagga tccagcgtgc ccggctcgc 960
agaagggcg ccgctgcctc ggcacatttg gcgcacacc gcgcacgcgc gtcctgaggg 1020
cgcggcgcct cgctccgacc gcggacggttc ctgggggtctg acacgcgcgc gcccacgac 1080
acgcctaccc ccggccggcc gcgcctgcct ccgcagcgcgc ccggcagaga 1140
atccgctag ccgctccctc ccgcagcgcgc ccgcagcgcgc ccggcagaga 1200
cggagcacc gcggagctgc caggtctgcc tggagccacat tgcagccgccc ccgcggtcgc 1260
gggctggg ctcgcacc cgcgtg 1287
<210> SEQ ID NO 9
<211> LENGTH: 1287
<212> TYPE: DNA
<213> ORGANISM: Kokuria rhizophila

<400> SEQUENCE: 9
atgacagagcc gttcggcga gacccgtagc gaaacagggcc cgacgctgct ggcgtccggtg  60
tacctctgcc cagcgcgctg cgcggcgcat gcgcgtga gctgctggca  120
gttgcatgc cgtctgctgg taaacaaaaat gttctgggtg gcggcagaa ggcggctaaa  180
cgtctctatg ataccagcgc ctggctgctg ggcggcgcg gacgcgggca gcgcgggca  240
cgtctgttgg gcgcaggggc gttcagagct cgtggcaggg gacgcgggca  300
aaccacacttgg gcgatagttg ttaggaaatg gttggagcgt tggacagccg ttaaagagc  360
gcgccgctgc gcggagacgc gcgcagagag gcggagacgc gcgcagagag gcgcagagag  420
ccgacgacagcg ccggagacgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag  480
cggcgagacgc cgccagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  540
ccgacgagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  600
caccagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  660
gcgccgagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  720
ggctgctgctgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  780
ggcgcgcgcgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  840
tgcctgctgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  900
tgcctgctgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag  960
ccggcgcgcgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1020
ggcgcgcgcgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1080
accccttcg ccgcgggcgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1140
tgcctgctgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1200
cgcgggagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1260
ggcgcgcgcgc gcgcagagag gcgcagagag gcgcagagag gcgcagagag gcgcagagag 1320

<210> SEQ ID NO 10
<211> LENGTH: 428
<212> TYPE: PRT
<213> ORGANISM: Kokuria rhizophila

<400> SEQUENCE: 10
Met Thr Ser Pro Phe Gly Glu Thr Ser Pro Phe Gly Glu Gly Pro Ser Leu  1    5     10     15
Leu Arg Ser Gly Tyr Leu Phe Ala Ser Arg Ala Arg Arg Arg Ala Gly  20   25   30
Leu Ser Ser Asp Ser Gly Cys Pro Val Arg Met Pro Leu Leu Gly Lys  35   40   45
Gln Thr Val Leu Val Arg Gly Glu Gly Val Lys Leu Phe Tyr Asp  50   55   60
Thr Ser Arg Val Arg Arg Asp Gly Ala Met Pro Gly Val Val Gln Gly  65   70   75   80
-continued

Pro Leu Phe Gly Ala Gly Ala Val His Gly Leu Asp Gly Glu Ala His
85 90 95
Arg Val Arg Lys Asn Gln Leu Ala Asp Met Ala Tyr Glu Asp Glu Arg
100 105 110
Val Ala Ala Tyr Lys Pro Phe Val Ala Glu Leu Glu Asn Leu Val
115 120 125
Ala Arg Trp Lys Asp Gly Asp Asn Val Tyr Asp Ser Thr Ala Ile Ala
130 135 140
Phe Gly Arg Ala Ser Phe Arg Trp Ala Gly Leu Gln Trp Gly Val Pro
145 150 155 160
Glu Met Asp Arg Trp Ala Arg Met Ser Arg Leu Asp Thr Phe
165 170 175
Gly Arg Pro Ala Thr His Leu Val Ser Arg Leu Asp Arg Ile Ala Leu
180 185 190
Asp Arg Arg Phe Ala Ala Leu Ile Lys Asp Val Arg Ala Gly Lys Val
195 200 205
Asn Ala Pro Glu Asp Ser Val Leu Ala His Met Ala Ala Leu Val Asp
210 215 220
Glu His Gly Glu Leu Val Asp Ala Lys Thr Ala Gly Ile Glu Leu Gln
225 230 235 240
Asn Leu Thr Arg Pro Asn Val Ala Val Ala Arg Phe Ala Ala Phe Ala
245 250 255
Ala Thr Ala Leu Val Glu His Pro Glu Trp Val Glu Arg Ile Arg Ala
260 265 270
Ala Ser Glu Gin Arg Gly Gly Thr Leu Leu Asp Val Pro Glu Ala Val
275 280 285
Ala Phe Ala Gin Glu Val Arg Asp Val Tyr Pro Phe Val Pro Met Leu
290 295 300
Pro Ala Glu Val Thr Gin Asp Thr Glu Ile Gin Gly Cys Pro Val His
305 310 315 320
Lys Gly Glu Arg Val Leu Asp Ile Leu Gly Thr Asn Thr Asp Pro
325 330 335
Thr Ser Trp Asp Arg Ala Ala Thr Phe Asp Pro Glu Arg Phe Leu Gly
340 345 350
Val Glu Asp Ala Glu Ala Ile Thr Thr Phe Ile Pro Glu Gin Gly Gly Ala
355 360 365
Glu Val Arg Thr Gly His Arg Cys Pro Gly Gly Lys Ile Ala Val Thr
370 375 380
Ser Leu Ser Ala Ala Val Ala Leu Cys Arg Pro Glu Val Gin Leu
385 390 395 400
Pro Gly Asp Gin Asp Leu Thr Phe Ser Trp Thr His Met Leu Thr
405 410 415
Arg Pro Val Thr Gly Val Arg Val Arg Thr Thr Arg
420 425

<210> SEQ ID NO 11
<211> LENGTH: 1275
<212> TYPE: DNA
<213> ORGANISM: Methylobacterium populi
<400> SEQUENCE: 11
atgcgcggctg ccatgccac ccacccyytc cgcaasagcag gcaccctgcc gcgtggcga
gtccagata gcacgtgcc gcctgccgc gcaggttaac gcctcatccg taacgcgttg 120
cggcgtcag accagacact gttgcaacgc gtgcttgctg tcagccgcgtt ccatctgatt 180
tcttgccgct gccggcccc ccaacctttac gcggcctacc gcgctgccct gcgacctgca 240
cgtgccgga gccgggggct tcgtgatcaaa gacacgcgaag cgctgttgcct ccgagcgctg 300
gccgacaccc tggcaaggtt ggtgctgttg tcgctgctg tgggtgcttg tgccttggcc 360
cgttgaggtgc gcctgagcct gcgctgcgct gcgcagcgcgc gcggcgcgtc ctgctgcttg 420
gataggttg gttctgtcga ccgagcact gcggctgcga ccgccgcgct gcgctgcttg 480
gttgggttgc gcgctgccgg gccgaagctg gatgctgctg cggctgcgct gcgctgcgcc 540
attgcctgca cctggctgct ggtgctggcc gcagccgctg gtccgtcgtg aggccgcggc 600
acagagcgtt gcgctgccgg gcctgctgct gcgggctgcg gcgctgcttg gcgctgcttg 660
gcggtgctgc cctggccact ccgccagcct gggcagcact gcgctgcgct gcgctgcttg 720
acagctgccgtg gcctgctgct gcctgccgca cctggctgct gcgctgcttg gcgctgcttg 780
attctgttgc gctggctgct gcctgccgca cctggctgct gcgctgcttg gcgctgcttg 840
gtggctggag cttttaccg gaagtgctgtg tggctgccgt gcgctgcttg gcgctgcttg 900
ctttcctgg gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 960
gataggttgc gcgtttgctg tggccgtcgt ctttttggcg gctgacgttt gcggcgggca 1020
gaaagctgcc ggcgtgctgg ccgccagcct ggtgctggcc gcagccgctg gtccgtcgtg 1080
ctggcgcgg gcggtggctg gcgctgctgg gcggtggctg tcgctgctg gcgctgctgct 1140
cggctgcga gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc 1200
cggctgcga gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc 1260
gtgcgccgct gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc gcgccgcgcgc 1275

<210> SEQ ID NO 12
<211> LENGTH: 424
<212> TYPE: PRT
<213> ORGANISM: Methylobacterium populi
<400> SEQUENCE: 12

Met Pro Ala Ala Ile Ala Thr His Arg Phe Arg Lys Ala Arg Thr Leu
  1    5   10   15
Pro Arg Glu Pro Ala Pro Asp Ser Thr Leu Ala Leu Leu Arg Glu Gly
  20  25   30
Tyr Gly Phe Ile Arg Asn Arg Cys Arg Arg His Arg Ser Asp Leu Phe
  35  40   45
Ala Ala Arg Leu Leu Leu Ser Pro Val Ile Cys Met Ser Gly Ala Glu
  50  55   60
Ala Ala Arg His Phe Tyr Asp Gly His Arg Phe Thr Arg Arg His Ala
  65  70   75   80
Leu Pro Pro Thr Ser Phe Ala Leu Ile Gln Asp His Gly Ser Val Met
  85  90   95
Val Leu Asp Gly Ala Ala His Leu Ala Arg Lys Ala Met Phe Leu Ser
100 105 110
Leu Val Gly Glu Ala Ala Leu Gin Arg Leu Ala Gly Leu Ala Glu Arg
115 120 125
His Trp Arg Glu Ala Val Ser Gly Trp Ala Arg Lys Asp Thr Val Val
130 135 140
<table>
<thead>
<tr>
<th>Residue</th>
<th>Identity</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Asp</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Val</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Phe</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Arg</td>
<td>Gly</td>
<td>His</td>
</tr>
<tr>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>Ile</td>
<td>Asp</td>
<td>Glu</td>
</tr>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Arg</td>
<td>Thr</td>
<td>Ile</td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Asn</td>
<td>Ala</td>
<td>Arg</td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>His</td>
<td>Gln</td>
<td>Arg</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Thr</td>
<td>Asp</td>
<td>Glu</td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Arg</td>
<td>Val</td>
<td>Arg</td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Glu</td>
<td>Trp</td>
<td>Val</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Trp</td>
<td>His</td>
<td>Glu</td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Ile</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Gly</td>
<td>His</td>
<td>Arg</td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Arg</td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Leu</td>
<td>Thr</td>
<td>Leu</td>
</tr>
<tr>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Val</td>
<td>Met</td>
<td>Arg</td>
</tr>
<tr>
<td>420</td>
<td>430</td>
<td>435</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 13
<211> LENGTH: 1244
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<400> SEQUENCE: 13

atgaatgagc agatcaca tcacaaagtt ctgcataaca gtcgacact gctgaggas 60
gggttatttt ttatasaasaa cagacacagag cgctacaatt cagatcgttt tcagcccgt 120
tttgtgagaa aaaaaatcta ttcagcact ggcgtgtagg cggcgaggt gtttgattatag 180
aaggtagat tcagcagcga gaagcgtcgg gctaacggag tcgagaaactc gctgtgagttgg 240
gttaatgtga ttcagaggtgt gatgagcagc ggcgtatatc atcaggaagat gtttttctcg 300
tcagagatg cagccgcgca cacaacagtt tgctgagttg ttggtgagtt gcagcagagaa 360
gcagcgatca caagatggga gaagccagat gagtttggtgt tatgtgaaga agccaaagaa 420
atcctgtgcc ggtagcgtg ctatgggcc ggtgttcggt tgaagggacc ggaagtcaaa 480
gaagagcgcc atgcctctg tcagtggtgc gacgcgttgc gtcgctggcg acgccggcat 540
tggagaagga gagaaggaag ggcggcttgg gcagagtgga ttgaaggtct gattgaagat 600
gctcggcgcg ctgtggagaa agcaaccttc ccgaacgcgag tcgatgaagat ggcggcattc 660
acacaagaag atggagaagcc gttgggtcag cgcagggcag ccctagtgcg cattaatgta 720
tgctggtccta ttttcatattc ttcgtggttt cagctttgagc gttctcatgaag 780
cattcagaga atcaggaattc gttcgcgttct ggaaacagcc gggagagaag aatgtttgtg 840
caggaggtcgg cgaatattta tgcgacctgcgg ctgtttttag ggcgcctggt gcagaaagat 900
tttagtatgca ataatcttgca gtttaaagag ggccactcgg tgcgtctgtta tttatgaga 960
acgaacacg accctcgtgc acgctagtc atcagctgat cccgtagat ctcgccggag aaggtttcgc 1020
gagagggaga aacatctgtt tgtatgagt ctctcaaggg ggccgccagc cgagaagagc 1080
cacgcgtcct caggagggag cattacaatg gaagctatgaa aagcagagct ggtacctc 1140
gttcatagag tgtgctggaa cttcctctgc attacagtct ccgcaagagt 1200
cctattgcg ctaaagggcg ctctgtatag aagccgaactc gcagaaagat 1254

<210> SEQ ID NO: 14
<211> LENGTH: 417
<212> TYPE: PRT
<213> ORGANISM: Bacillus subtilis
<400> SEQUENCE: 14
Met Asn Glu Gln Ile Pro His Asp Lys Ser Leu Asp Asn Ser Leu Thr
1 5 10 15
Leu Leu Lys Glu Gly Tyr Leu Phe Ile Lys Asn Arg Thr Glu Arg Tyr
20 25 30
Asn Ser Asp Leu Phe Glu Ala Arg Leu Leu Gly Lys Asn Phe Ile Cys
35 40 45
Met Thr Gly Ala Glu Ala Lys Val Phe Tyr Asp Thr Asp Arg Phe
50 55 60
Gln Arg Gln Asp Ala Leu Pro Lys Arg Val Gln Lys Ser Leu Phe Gly
65 70 75 80
Val Asn Ala Ile Gln Gly Met Asp Gly Ser Ala His Ile His Arg Lys
85 90 95
Met Leu Phe Leu Ser Leu Met Thr Pro Pro His Gln Lys Arg Leu Ala
100 105 110
Glu Leu Met Thr Glu Glu Trp Lys Ala Ala Val Thr Arg Trp Glu Lys
115 120 125
Ala Asp Glu Val Val Leu Phe Glu Ala Lys Gly Ile Leu Cys Arg
130 135 140
Val Ala Cys Tyr Trp Ala Gly Val Pro Leu Lys Glu Thr Glu Val Lys
145 150 155 160
Glu Arg Ala Asp Asp Phe Ile Asp Met Val Asp Ala Phe Gly Ala Val
165 170 175
Gly Pro Arg His Trp Lys Gly Arg Ala Arg Ile Gly Ala Glu Glu
180 185 190
Trp Ile Glu Val Met Ile Glu Asp Ala Arg Ala Gly Leu Leu Lys Thr
195 200 205
1. A microbial host cell engineered to produce an olefin, comprising a heterologous polynucleotide sequence encoding a polypeptide having fatty acid decarboxylase activity.

212. The recombinant microbial host cell of claim 211, wherein the recombinant microbial host cell has a sequence encoding an acyl-CoA synthase that is homologous to a corresponding wild-type microbial host cell.

213. The recombinant microbial host cell of claim 212, wherein the recombinant microbial host cell expresses an increased level of the polypeptide having fatty acid decarboxylase activity relative to a corresponding wild-type microbial host cell.

214. The recombinant microbial host cell of claim 212, wherein a gene encoding an acyl-CoA synthase is attenuated or deleted in the recombinant microbial host cell such that the reduced level of acyl-CoA synthase relative to the corresponding wild-type microbial host cell.

215. The recombinant microbial host cell of claim 214, wherein the gene encoding an acyl-CoA synthase isfadD.

216. The recombinant microbial host cell of claim 212, wherein the recombinant microbial host cell is engineered to express, overexpress, or attenuate a gene encoding a thioesterase relative to a corresponding wild-type microbial host cell.

217. The recombinant microbial host cell of claim 210, wherein the recombinant microbial host cell is selected from the group consisting of Escherichia Coli, Bacillus subtilis, and Saccharomyces cerevisiae.

218. The recombinant microbial host cell of claim 217, wherein the recombinant microbial host cell is Escherichia Coli.

219. A cell culture comprising the recombinant microbial host cell of claim 210, wherein an olefin is produced in the cell culture when the recombinant microbial host cell is cultured in a medium containing a substrate under conditions effective to express the polypeptide having fatty acid decarboxylase activity.

220. The cell culture of claim 219, wherein the olefin is a terminal olefin.

221. The cell culture of claim 220, wherein the terminal olefin is selected from the group consisting of a C4-C25 terminal olefin, a C13-C21 terminal olefin, nonadecane, methylpentadecane, 2-ethylhexadecane, 3-methylheptadecane, and pentadecane, an unsaturated terminal olefin, a monounsaturated terminal olefin, a straight chain terminal olefin, a branched chain terminal olefin, and a terminal olefin having a cyclic moiety.

222. The cell culture of claim 219, wherein the olefin is a terminal olefin.

-continued
The cell culture of claim 219, wherein the substrate is a carbon source.

The cell culture of claim 223, wherein the carbon source is a fatty acid.

The cell culture of claim 224, wherein the fatty acid is selected from the group consisting of a C₆-C₉ fatty acid, a C₁₄-C₂₀ fatty acid, tetradecanoic acid, hexadecanoic acid, octadecanoic acid, behenic acid, methyl eicosanoic acid, eicosanoic acid, methyl octadecanoic acid, stearic acid, and palmitic acid, an unsaturated fatty acid, a monounsaturated fatty acid, a saturated fatty acid, a straight chain fatty acid, and a cyclic moiety.

The cell culture of claim 226, wherein the carbon source is a carbohydrate.

The cell culture of claim 226, wherein the carbohydrate is glucose.

A method of producing an olefin comprising culturing the bacterial host cell of claim 210, under conditions sufficient to allow expression of the polypeptide having fatty acid dehydrogenase activity.

The method of claim 228, further comprising culturing the microbial host cell in a culture medium in the presence of at least one biological substrate for the polypeptide having fatty acid dehydrogenase activity.

The method of claim 229, further comprising isolating the olefin from the culture medium.

An olefin produced by the method of claim 230, wherein the olefin has a δ¹³C of about −15.4 or greater, about −15.4 to about −10.9, or about −13.92 to about −13.84.

An olefin produced by the method of claim 230, wherein the olefin has an f₁₈₀ of at least about 1.003, at least about 1.01, at least about 1.5, or about 1.111 to about 1.124.