



US 20250263731A1

(19) United States

(12) Patent Application Publication

Kruis et al.

(10) Pub. No.: US 2025/0263731 A1

(43) Pub. Date: Aug. 21, 2025

(54) GENETICALLY MODIFIED
METHYLOTROPHIC BACTERIA
PRODUCING LACTATE*CI2N 1/32* (2006.01)
CI2N 9/04 (2006.01)
CI2N 9/10 (2006.01)
CI2N 9/12 (2006.01)
CI2P 7/56 (2006.01)

(71) Applicant: Acies Bio d.o.o., Ljubljana (SI)

(52) U.S. Cl.
CPC *CI2N 15/74* (2013.01); *CI2N 1/20*
(2013.01); *CI2N 1/32* (2013.01); *CI2N*
9/0006 (2013.01); *CI2N 9/1029* (2013.01);
CI2N 9/1229 (2013.01); *CI2P 7/56* (2013.01);
CI2Y 101/01027 (2013.01); *CI2Y 203/01184*
(2013.01); *CI2Y 207/04001* (2013.01)

(21) Appl. No.: 18/858,137

(57) ABSTRACT

(22) PCT Filed: Apr. 21, 2023

(86) PCT No.: PCT/EP2023/060503

§ 371 (c)(1),
(2) Date: Oct. 18, 2024

(30) Foreign Application Priority Data

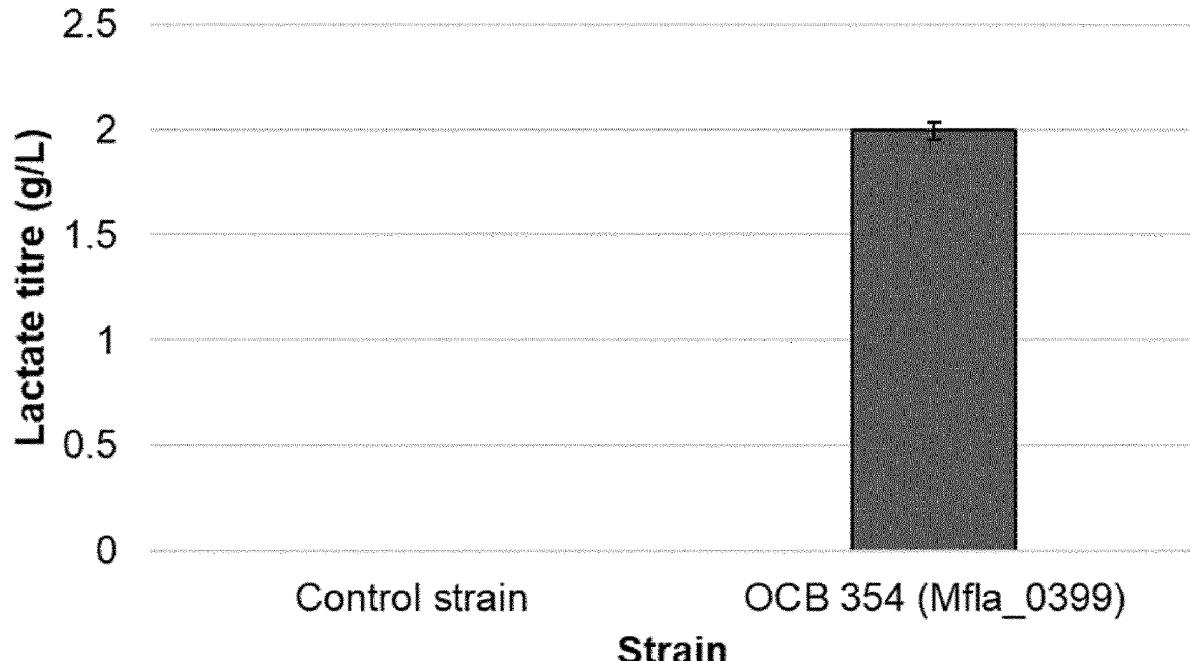
Apr. 21, 2022 (EP) 22169245.2

The present invention generally relates to the biotechnology engineering, and specifically to genetically modified methylovrophic bacteria which produce lactate. More specifically, the present invention provides a methylovrophic bacterium modified to have an increased expression of a polypeptide having lactate dehydrogenase activity. The present invention further provides a method for producing lactate using a genetically modified bacterium of the present invention.

Specification includes a Sequence Listing.

(51) Int. Cl.
CI2N 15/74 (2006.01)
CI2N 1/20 (2006.01)

Lactate production by OCB 354 expressing Mfla_0399 in shake flasks



Lactate production by OCB 354 expressing Mfla_0399 in shake flasks

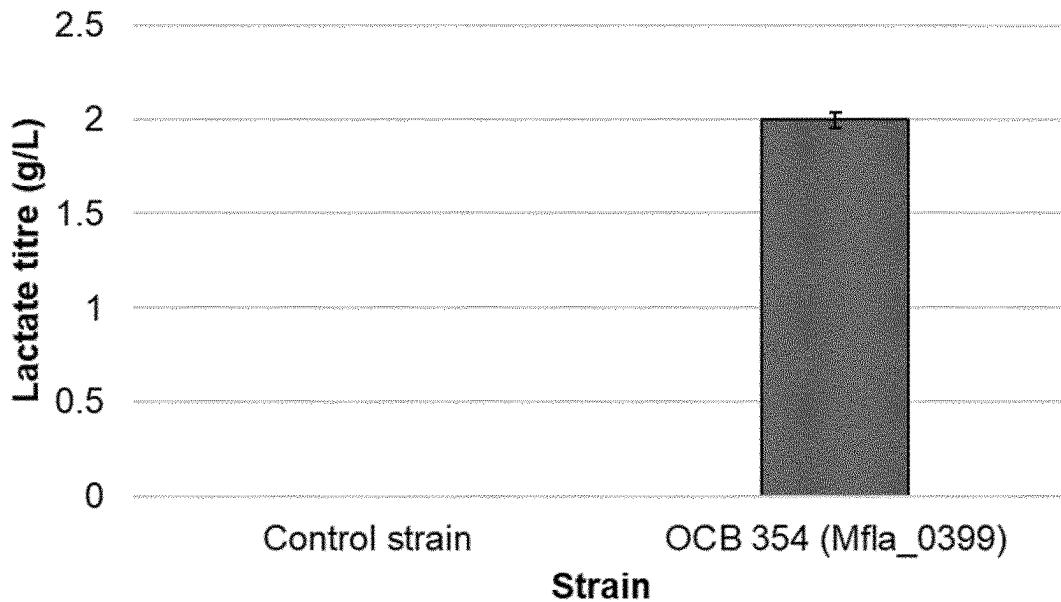


Figure 1

L-Lactate production from methanol by a strain expressing L-LDH

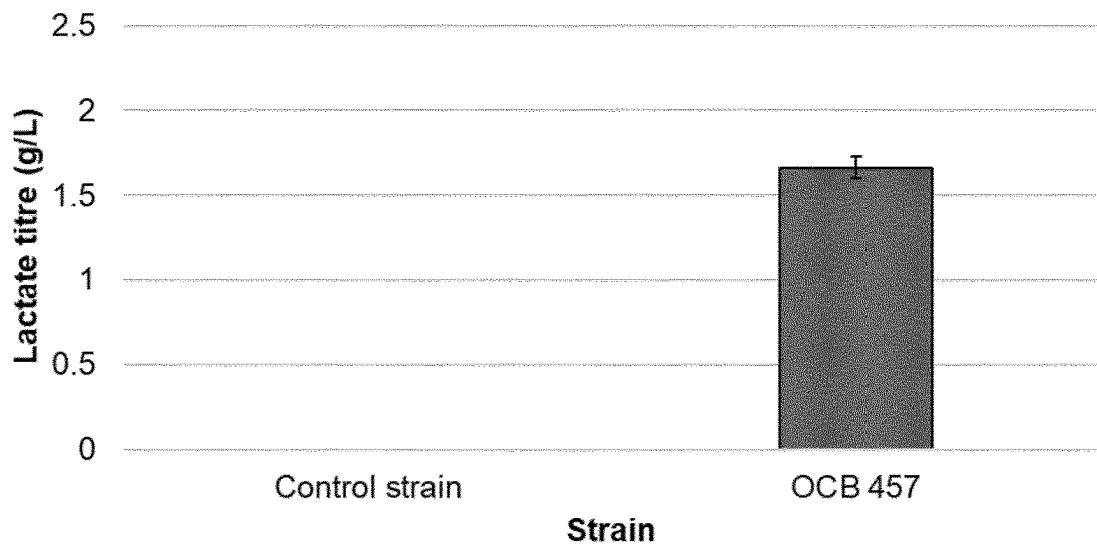


Figure 2

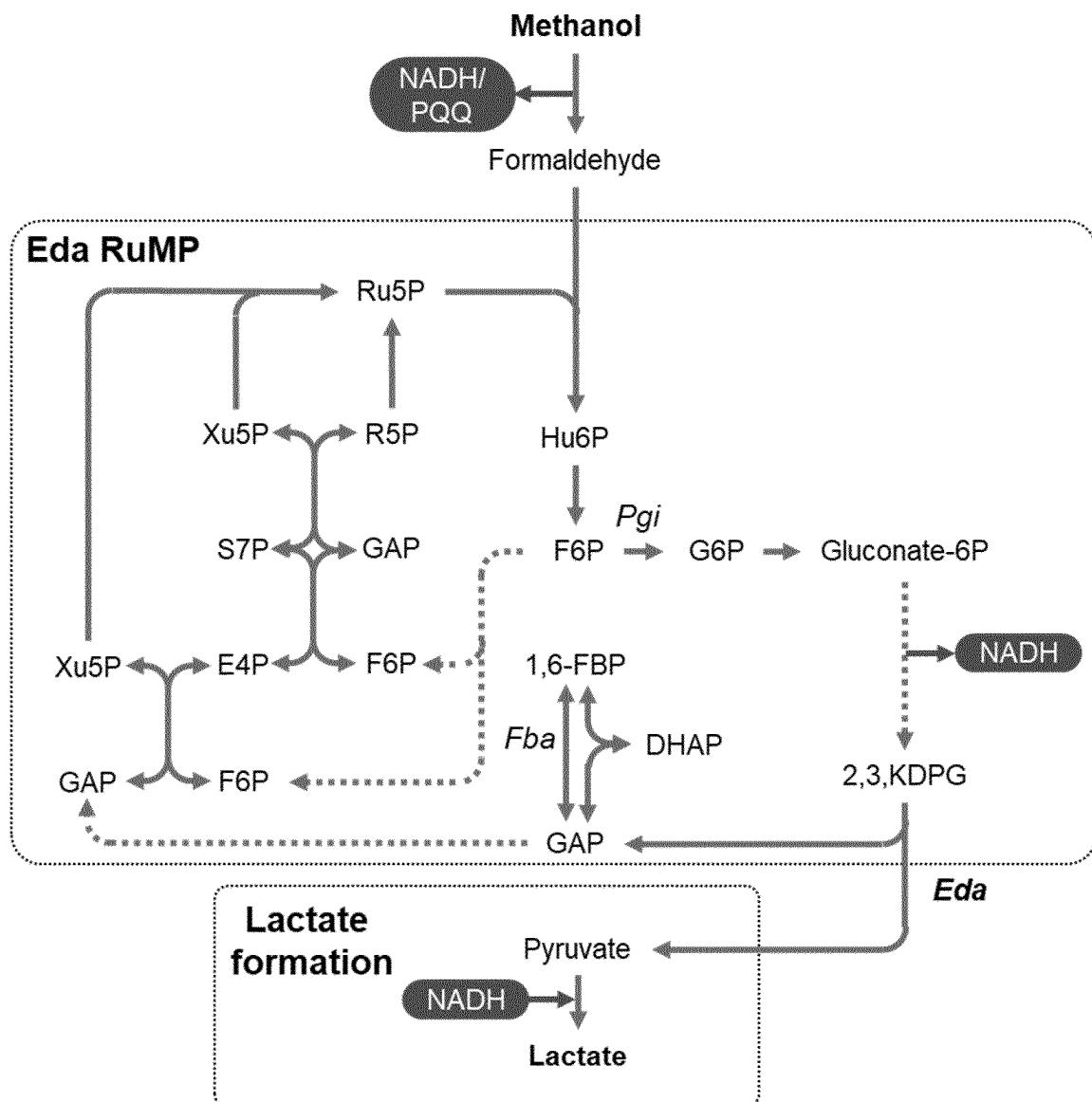


Figure 3

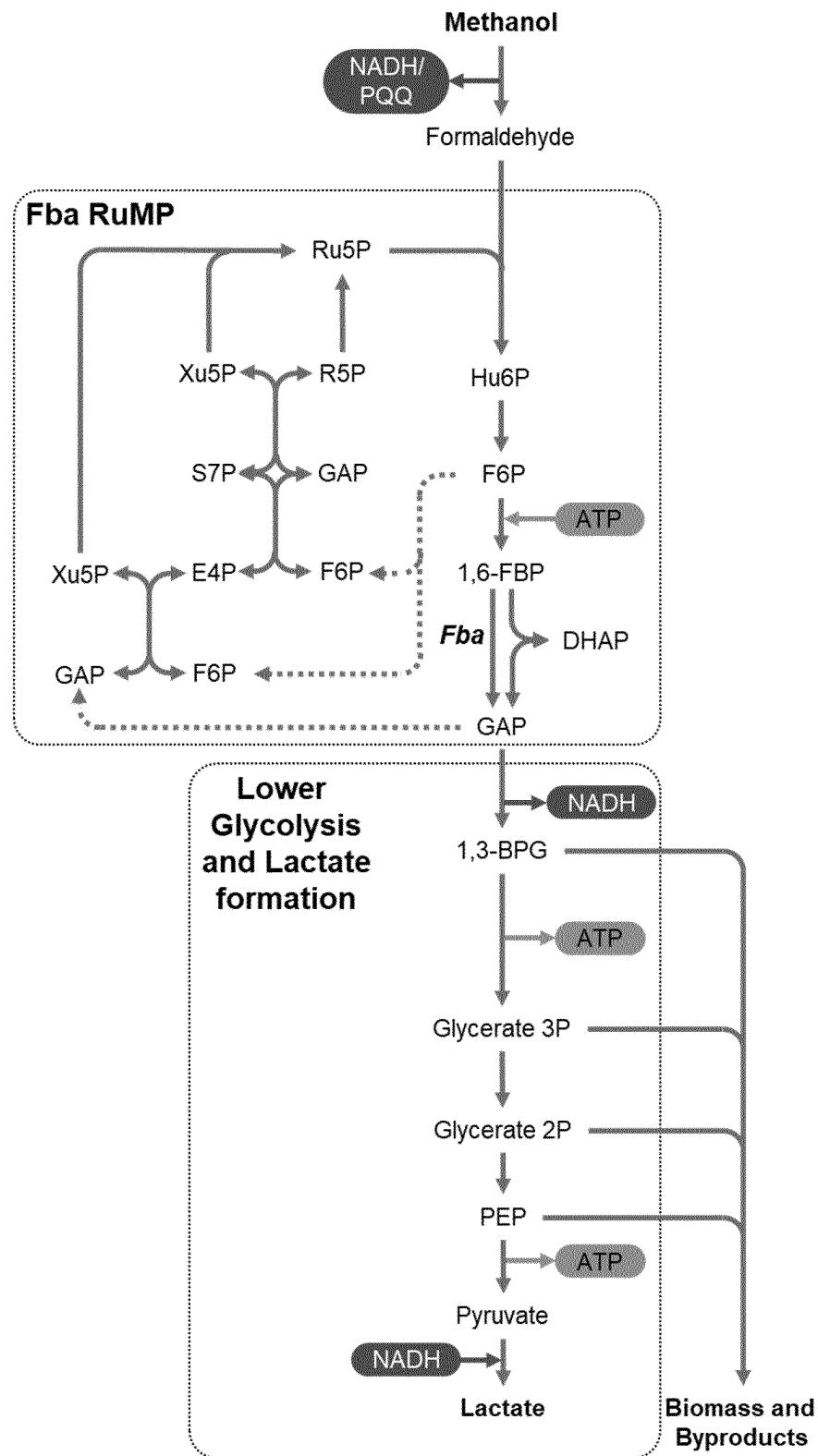


Figure 4

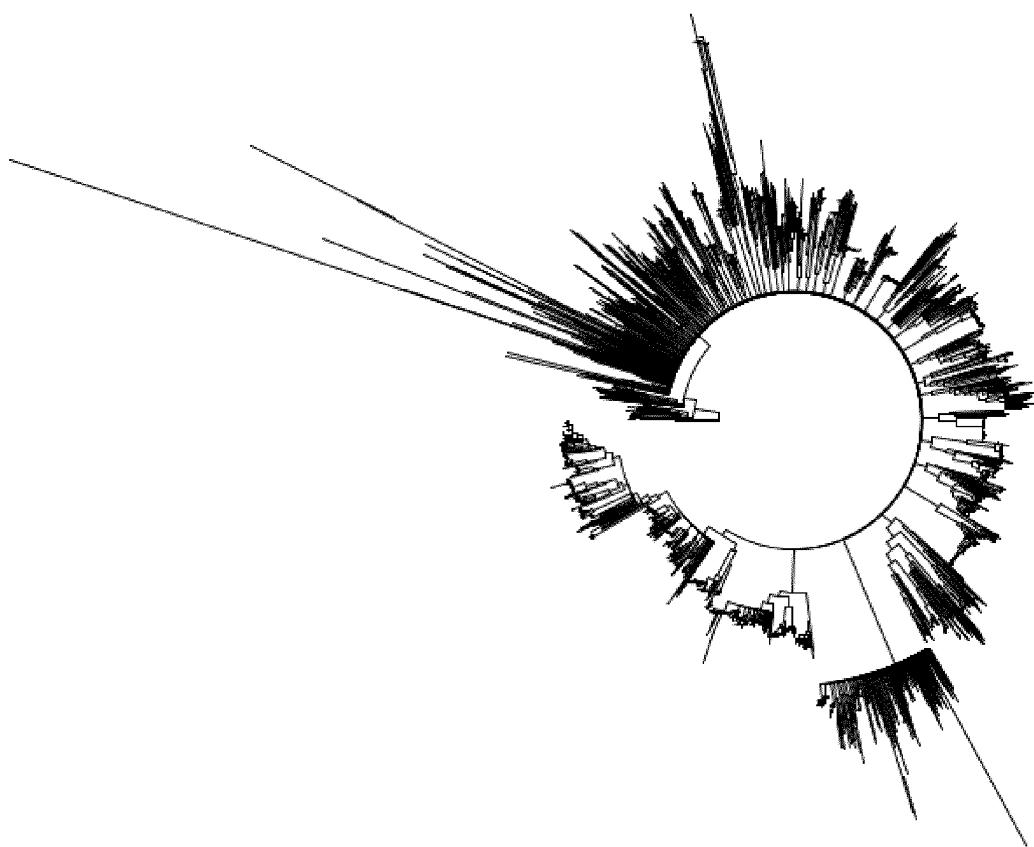


Figure 5

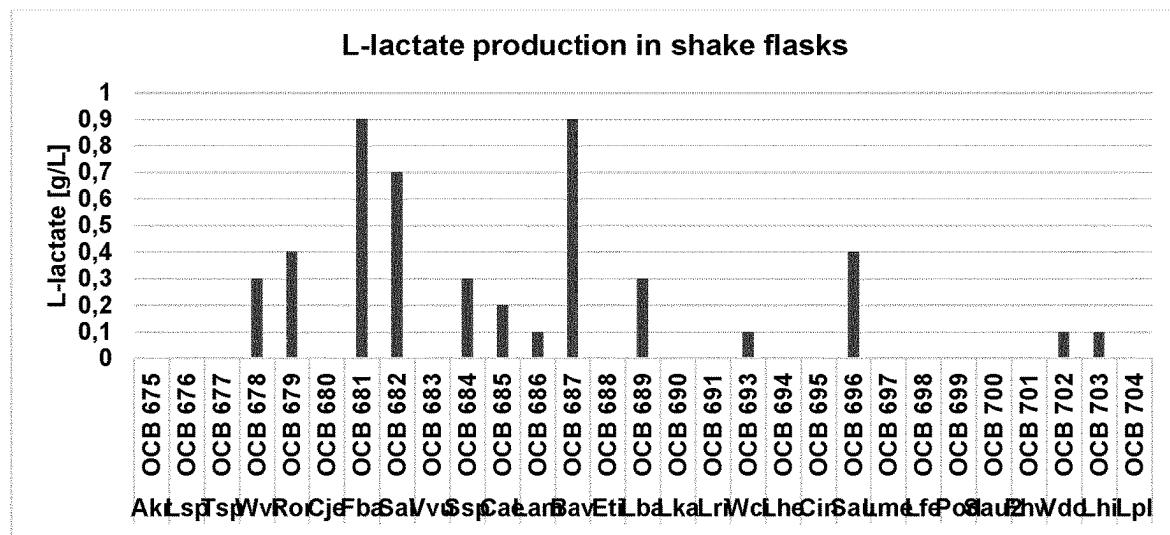


Figure 6

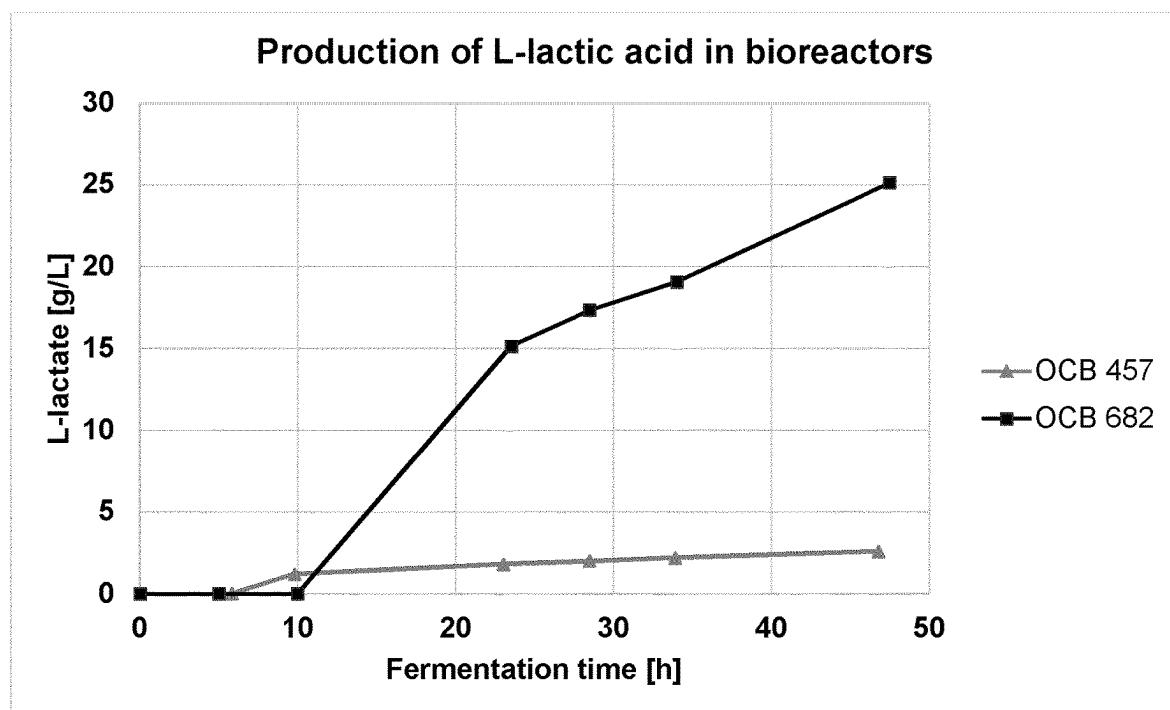


Figure 7

GENETICALLY MODIFIED METHYLOTROPHIC BACTERIA PRODUCING LACTATE

TECHNICAL FIELD OF THE INVENTION

[0001] The present invention generally relates to the biotechnology engineering, and specifically to genetically modified methylotrophic bacteria which produce lactate. More specifically, the present invention provides a methylotrophic bacterium modified to have an increased expression of a polypeptide having lactate dehydrogenase activity. The present invention further provides a method for producing lactate using a genetically modified bacterium of the present invention.

BACKGROUND OF THE INVENTION

[0002] Lactic acid is a widely used building block for biodegradable polymer production. It is produced commercially by fermentation of carbohydrates by various bacteria, such as *Bacillus coagulans*. The carbohydrates, in particular C6 sugars, such as glucose are converted to lactate through an anaerobic or severely-oxygen limited process. The glucose is first imported into the host cell and metabolised to pyruvate, which is subsequently converted to lactate by the enzyme lactate dehydrogenase.

[0003] During fermentation, high lactate concentrations are produced, which is toxic to the production organism. To circumvent this problem, calcium hydroxide or calcium carbonate is used to regulate pH during the fermentation, which results in the formation of Ca-lactate salts. These are insoluble and precipitate during the fermentation, reducing the toxicity of the lactate. At the end of the process, the Ca-lactate is acidified with H₂SO₄ to convert it to lactic acid, which is then further purified to the final product, producing gypsum (CaSO₄) as a by-product.

[0004] Two enantiomers of lactate can be produced by microorganisms, D-lactate, and L-lactate. The stereospecificity of the lactate is determined by the lactate dehydrogenase that converts pyruvate to lactate and is generally organism specific. By replacing an L-lactate producing LDH with a D-lactate producing LDH through genetic methods, the final lactate form produced by the host can be controlled. L-lactic acid is currently the main commercially produced lactate form. Both lactic acid enantiomers can be polymerized into polylactic acid (PLA) to form L-PLA and D-PLA, which can be mixed subsequently to form DL-PLA.

[0005] Bio-based products, such as lactic acid, can be produced from various substrates, such as carbohydrates and methanol. Carbohydrates have been the dominant substrate used for microbial chemical production due to their widespread availability. While carbohydrates may be a sustainable substrate in terms of atmospheric CO₂ release, they compete with food and feed chains. Because of the growing global population, carbohydrate-derived substrates will become less and less sustainable in the long term. On the other hand, bio-based methanol production from wood biomass or CO₂-sequestration does not interfere with food and feed production.

[0006] Bacteria that can utilize methanol for growth and product formation (methylotrophs) have been identified. Examples include *Bacillus methanolicus*, *Methylobacterium extorquens*, *Methylobacillus glycogenes* and *Methylobacillus flagellatus*. Methylotrophic bacteria can be divided into

two broad groups based on the underlying metabolism that they use to assimilate methanol. In both groups, methanol is first converted to formaldehyde using a methanol dehydrogenase. The first group, which includes *M. extorquens*, assimilate formaldehyde by reacting it with glycine, forming serine in the pathway known as the serine cycle.

[0007] The second group of methylotrophic bacteria use the Ribulose-monophosphate (RuMP) cycle to react formaldehyde with RuMP to form C6-compounds that are then metabolized further. *B. methanolicus* and *Methylobacilli* belong to the RuMP cycle group of methylotrophs. The RuMP pathway has two sub-variants, the Eda pathway and the Fba pathway. The FBA variant depends on the fructose bisphosphate aldolase (Fba) enzyme to cleave the C6 intermediates of the RuMP cycle. This generates two GAP moieties, one of which is used for RuMP regeneration, and the other exists the cycle and enters lower glycolysis. The Eda dependent RuMP cycle uses the 2-keto-3-deoxy glutonate-6P aldolase (2,3-KDPG aldolase, or Eda for Entner-Doudorff pathway aldolase) to extract C3 compounds. The Eda enzyme splits the C6 compound 2-keto, 3-deoxy glutonate-6P (2,3-KDPG) into glyceraldehyde-3P (GAP) and pyruvate. GAP is used to replenish the RuMP cycle for the next cycle of methanol assimilation, while pyruvate exits the cycle.

[0008] Both groups of methylotrophic bacteria have been used to produce various biochemicals from methanol. However, lactate production with methylotrophs has not yet been reported. Many glucose-utilizing organisms such as *Escherichia coli* or lactic acid bacteria produce lactate under anaerobic conditions. In the absence of oxygen, lactate acts as the final acceptor of electrons released during glucose oxidation. This allows the organisms to produce ATP and prevents a complete halt in their metabolism. Oxygen limitation accompanied by overabundance of a carbon source can also lead to lactate production, in which case it is considered an “overflow” metabolite.

[0009] This is not possible in methylotrophs since they are obligate aerobes and require oxygen for methanol assimilation. Furthermore, methylotrophs have the option of generating ATP from methanol before it is assimilated through the RuMP or serine cycles via formaldehyde dissimilation pathways or during the RuMP cycle via the dissimilatory RuMP pathway. Both pathways generate NADH, which can enter the respiratory chain to produce ATP. Any carbon that is not needed for biomass formation is therefore converted into energy and CO₂ before it even enters the central carbon metabolism. It is therefore highly unexpected that methylotrophs would efficiently produce lactate as glucose-utilizing organisms do.

SUMMARY OF THE INVENTION

[0010] The object of the present invention is to provide means allowing efficient production of lactate at higher nominal yield. This is achieved by the present inventors who have engineered genetically modified methylotrophic bacteria which produce lactate.

[0011] More specifically, the present inventors have engineered methylotrophic bacterial strains, which have an increased expression of a polypeptide having lactate dehydrogenase activity. As shown in the Examples, such engineered bacterial strains surprisingly show unusually high titers of lactate in the supernatant.

[0012] The present invention this provides in a first aspect a methylotrophic bacterium which has been to have an increased protein expression of a polypeptide having lactate dehydrogenase activity compared to an otherwise identical bacterium that does not carry said modification.

[0013] The present invention further provides in a second aspect a method for producing lactate, comprising cultivating a bacterium according to the present invention under suitable culture conditions in a suitable culture medium.

[0014] The present invention may be further summarized by the following items:

[0015] 1. A genetically engineered methylotrophic bacterium which has been modified to have an increased protein expression of a polypeptide having lactate dehydrogenase activity compared to an otherwise identical bacterium that does not carry said modification.

[0016] 2. The bacterium according to item 1, wherein the increase in protein expression of the polypeptide having lactate dehydrogenase activity is achieved by increasing the number of copies of a nucleotide sequence encoding said polypeptide having lactate dehydrogenase activity.

[0017] 3. The bacterium according to item 2, wherein the increase in the number of copies of the nucleotide sequence encoding said polypeptide is achieved by introducing into the bacterium at least one exogenous nucleic acid molecules comprising at least one nucleotide sequence encoding said polypeptide having lactate dehydrogenase activity.

[0018] 4. The bacterium according to any one of items 1 to 3, wherein the bacterium comprises at least one exogenous nucleic acid molecule (such as a vector) comprising at least one nucleotide sequence encoding the polypeptide having lactate dehydrogenase activity.

[0019] 5. The bacterium according to item 3 or 4, wherein the exogenous nucleic acid molecule comprises at least one transcriptional unit comprising, from 5' to 3', a promoter that is functional in the bacterium to cause the production of an mRNA molecule and that is operably linked to a nucleotide sequence encoding said polypeptide having lactate dehydrogenase activity, and a transcriptional terminator sequence.

[0020] 6. The bacterium according to any one of items 3 to 5, wherein the exogenous nucleic acid molecule is a vector, such as a plasmid.

[0021] 7. The bacterium according to any one of items 3 to 5, wherein the exogenous nucleic acid molecule is stably integrated into the genome of the bacterium.

[0022] 8. The bacterium according to any one of items 1 to 7, wherein the increase in protein expression of the polypeptide having lactate dehydrogenase activity is achieved by modifying the ribosome binding site of an endogenous gene encoding the polypeptide having lactate dehydrogenase activity.

[0023] 9. The bacterium according to any one of items 1 to 8, wherein the increase in protein expression of the polypeptide having lactate dehydrogenase activity is achieved by increasing the strength of the promoter operably linked to an endogenous gene encoding the polypeptide having lactate dehydrogenase activity.

[0024] 10. The bacterium according to any one of items 1 to 9, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 1 to 130 and 182 to 212, and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least

about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 130 and 182 to 212.

[0025] 11. The bacterium according to any one of items 1 to 10, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 1 to 130 and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 130.

[0026] 12. The bacterium according to any one of items 1 to 11, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 1 to 48; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 48.

[0027] 13. The bacterium according to any one of items 1 to 12, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 12 to 48; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 12 to 48.

[0028] 14. The bacterium according to any one of items 1 to 12, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising the amino acid sequence of SEQ ID NO: 1; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of SEQ ID NO: 1.

[0029] 15. The bacterium according to any one of items 1 to 12, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 49 to 130; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 49 to 130.

[0030] 16. The bacterium according to any one of items 1 to 12, wherein the polypeptide having lactate dehydrogenase activity is not allosterically regulated by fructose-1,6-bisphosphate.

[0031] 17. The bacterium according to any one of items 1 to 12 and 16, wherein the polypeptide having lactate dehy-

drogenase activity has an amino acid not being histidine at the amino acid at position corresponding to *L. casei* LDH Histidine 188 as determined by sequence alignment.

[0032] 18. The bacterium according to any one of items 16 and 17, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 182 to 210; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 182 to 210.

[0033] 19. The bacterium according to any one of items 16 to 18, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 185, 186, 188, 189, 191, 194, 196 and 202; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 185, 186, 188, 189, 191, 194, 196 and 202.

[0034] 20. The bacterium according to any one of items 16 to 19, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 188, 189 and 194; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 188, 189 and 194.

[0035] 21. The bacterium according to any one of items 1 to 20, wherein the polypeptide having lactate dehydrogenase activity is a heterologous polypeptide having lactate dehydrogenase activity.

[0036] 22. The bacterium according to any one of items 1 to 21, which has been further modified to have a decreased expression and/or activity of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

[0037] 23. The bacterium according to any one of items 1 to 22, which has been modified to have a decreased expression of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

[0038] 24. The bacterium according to item 22 or 23, wherein the expression level of the endogenous polypeptide having polyphosphate kinase activity is decreased by at least 50%, such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.

[0039] 25. The bacterium according to any one of item 22 to 24, wherein the endogenous gene encoding said polypeptide having polyphosphate kinase activity has been inactivated.

[0040] 26. The bacterium according to item 25, wherein the endogenous gene encoding said polypeptide having polyphosphate kinase activity has been inactivated by deletion of part of or the entire gene sequence.

[0041] 27. The bacterium according to item 25 or 26, wherein the endogenous gene encoding said polypeptide having polyphosphate kinase activity has been inactivated by introducing or expressing in the bacterium a rare-cutting endonuclease able to selectively inactivating by DNA cleavage the endogenous gene encoding said polypeptide.

[0042] 28. The bacterium according to item 27, wherein said rare-cutting endonuclease is a transcription activator-like effector (TALE) nuclease, meganuclease, zinc-finger nuclease (ZFN), or RNA-guided endonuclease.

[0043] 29. The bacterium according to item 28, wherein the RNA-guided endonuclease is a catalytically inactive Cas9 protein.

[0044] 30. The bacterium according to item 29, which comprises (e.g., expresses) a single guide RNA (sgRNA) specifically hybridizing (e.g. binding) under cellular conditions with the genomic DNA encoding said polypeptide.

[0045] 31. The bacterium according to any one of items 22 to 24, wherein the expression of said endogenous polypeptide having polyphosphate kinase activity is decreased (e.g., inhibited) by transcriptional and/or translational repression of the endogenous gene encoding said polypeptide.

[0046] 32. The bacterium according to any one of items 22 to 24, wherein the expression of said endogenous polypeptide having polyphosphate kinase activity is decreased (e.g. inhibited) by introducing or expressing in the bacterium an inhibitory nucleic acid molecule that specifically hybridizes (e.g. binds) under cellular conditions with cellular mRNA and/or genomic DNA encoding said polypeptide.

[0047] 33. The bacterium according to item 32, wherein the inhibitory nucleic acid molecule is an antisense oligonucleotide, ribozyme or interfering RNA (RNAi) molecule.

[0048] 34. The bacterium according to item 33, wherein the interfering RNA molecule is a micro RNA (miRNA), small interfering RNA (siRNA) or short hairpin RNA (shRNA).

[0049] 35. The bacterium according to any one of items 1 to 22, which has been modified to have a decreased activity of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical microorganism that does not carry said modification.

[0050] 36. The bacterium according to item 35, wherein the activity of said polypeptide is decreased by at least one active-site mutation resulting in the reduction or loss of activity.

[0051] 37. The bacterium according to item 36, wherein the at least one active-site mutation is a non-conservative amino acid substitution.

[0052] 38. The bacterium according to any one of items 1 to 37, which has been further modified to have a decreased expression and/or activity of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

[0053] 39. The bacterium according to any one of items 1 to 38, which has been modified to have a decreased expression of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

- [0054] 40. The bacterium according to item 38 or 39, wherein the expression level of the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased by at least 50%, such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.
- [0055] 41. The bacterium according to any one of item 38 to 40, wherein the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity has been inactivated.
- [0056] 42. The bacterium according to item 41, wherein the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity has been inactivated by deletion of part of or the entire gene sequence.
- [0057] 43. The bacterium according to item 41 or 42, wherein the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity has been inactivated by introducing or expressing in the bacterium a rare-cutting endonuclease able to selectively inactivating by DNA cleavage the endogenous gene encoding said polypeptide.
- [0058] 44. The bacterium according to item 43, wherein said rare-cutting endonuclease is a transcription activator-like effector (TALE) nuclease, meganuclease, zinc-finger nuclease (ZFN), or RNA-guided endonuclease.
- [0059] 45. The bacterium according to item 44, wherein the RNA-guided endonuclease is a catalytically inactive Cas9 protein.
- [0060] 46. The bacterium according to item 45, which comprises (e.g., expresses) a single guide RNA (sgRNA) specifically hybridizing (e.g. binding) under cellular conditions with the genomic DNA encoding said polypeptide.
- [0061] 47. The bacterium according to item 38 or 39, wherein the expression of said endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased (e.g., inhibited) by transcriptional and/or translational repression of the endogenous gene encoding said polypeptide.
- [0062] 48. The bacterium according to item 38 or 39, wherein the expression of said endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased (e.g. inhibited) by introducing or expressing in the bacterium an inhibitory nucleic acid molecule that specifically hybridizes (e.g. binds) under cellular conditions with cellular mRNA and/or genomic DNA encoding said polypeptide.
- [0063] 49. The bacterium according to item 48, wherein the inhibitory nucleic acid molecule is an antisense oligonucleotide, ribozyme or interfering RNA (RNAi) molecule.
- [0064] 50. The bacterium according to item 49, wherein the interfering RNA molecule is a micro RNA (miRNA), small interfering RNA (siRNA) or short hairpin RNA (shRNA).
- [0065] 51. The bacterium according to item 38, which has been modified to have a decreased activity of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical microorganism that does not carry said modification.
- [0066] 52. The bacterium according to item 51, wherein the activity of said polypeptide is decreased by at least one active-site mutation resulting in the reduction or loss of activity.
- [0067] 53. The bacterium according to item 52, wherein the at least one active-site mutation is a non-conservative amino acid substitution.
- [0068] 54. The bacterium according to any one of items 1 to 53, which expresses a RuMP cycle.
- [0069] 55. The bacterium according to any one of items 1 to 54, which expresses an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle.
- [0070] 56. The bacterium according to any one of items 1 to 55, which expresses a polypeptide having phosphoglucoisomerase activity and a polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity.
- [0071] 57. The bacterium according to any one of items 1 to 56, wherein said bacterium belongs to the family Methylophilaceae or Methylbacteriaceae.
- [0072] 58. The bacterium according to any one of items 1 to 57, wherein said bacterium belongs to the genus *Methylobacillus*, *Methylobacterium* or *Methylocruberum*, preferably *Methylobacillus* or *Methylobacterium*.
- [0073] 59. The bacterium according to any one of items 1 to 58, which is selected from *Methylobacillus flagellatus*, *Methylobacillus glycogenes*, *Methylobacillus pratensis*, *Methylobacillus rhizosphaerae*, *Methylobacillus gramineus*, *Methylobacillus arboreus*, *Methylobacillus caricis*, *Methylobacillus methilvorans*, *Methylobacillus* sp, *Methylobacterium extorquens*, *Methylobacterium organophilum* and *Methylocruberum extorquens*.
- [0074] 60. The bacterium according to any one of items 1 to 58, wherein said bacterium is of the genus *Methylobacillus*.
- [0075] 61. The bacterium according to any one of items 1 to 58, which is selected from *Methylobacillus flagellatus*, *Methylobacillus glycogenes*, *Methylobacillus pratensis*, *Methylobacillus rhizosphaerae*, *Methylobacillus gramineus*, *Methylobacillus arboreus*, *Methylobacillus caricis*, *Methylobacillus methilvorans* and *Methylobacillus* sp.
- [0076] 62. The bacterium according to any one of items 1 to 58, wherein said bacterium is *Methylobacillus flagellatus*.
- [0077] 63. The bacterium according to any one of items 1 to 58, wherein said bacterium is *Methylobacillus glycogenes*.
- [0078] 64. Method for producing lactate comprising cultivating a bacterium as defined in any one of items 1 to 63 under suitable culture conditions.
- [0079] 65. Method for producing D-lactate comprising cultivating a bacterium as defined in item 13 under suitable culture conditions.
- [0080] 66. Method for producing L-lactate comprising cultivating a bacterium as defined in item 15 under suitable culture conditions.
- [0081] 67. Method for producing L-lactate comprising cultivating a bacterium as defined in any one of items 16 to 20.
- [0082] 68. The method according to any one of items 64 to 67, comprising cultivating said bacterium under suitable culture conditions in a culture medium comprising a reduced one-carbon compound, such as methanol, or a multi-carbon compound that contains no carbon-carbon bonds, such as dimethyl ether and dimethylamine.
- [0083] 69. The method according to item 68, wherein the culture medium comprises methanol.
- [0084] 70. The method according to any one of items 64 to 69, wherein the cultivation is performed in a bioreactor.

[0085] 71. Use of a bacterium as defined in any one of items 1 to 63 for producing lactate in a cultivation.

BRIEF DESCRIPTION OF THE FIGURES

[0086] FIG. 1: D-Lactate production from methanol by the strain OCB 354 that overexpressed the *Mfla_0399* gene (SEQ ID NO: 137, encoding SEQ ID NO: 1).

[0087] FIG. 2: L-Lactate production from methanol by the strain OCB 456 that overexpressed the *Pediococcus* *Idh* gene (SEQ ID NO: 149, encoding SEQ ID NO: 50)

[0088] FIG. 3: Eda variant of RuMP cycle for the generation of lactate

[0089] FIG. 4: Fba variant of RuMP cycle for the generation of lactate.

[0090] FIG. 5: Similarity tree of ~3700 L-LDH sequences without a Histidine at position corresponding to *L. Casei* His188

[0091] FIG. 6: Production of lactic acid in *M. flagellatus* strains expressing different L-LDH enzymes selected from the similarity tree

[0092] FIG. 7: Comparison of lactate production in *M. flagellatus* strains expressing *Pediococcus* L-LDH (OCB 457) and *Snodgrassella alvi* (OCB 682) L-LDH in an otherwise identical bioprocess

[0093] The present invention is now described in more detail below.

DETAILED DESCRIPTION OF THE INVENTION

[0094] Unless specifically defined herein, all technical and scientific terms used have the same meaning as commonly understood by a skilled artisan in the fields of biochemistry, genetics, and microbiology.

[0095] All methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, with suitable methods and materials being described herein. 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 prevail. Further, the materials, methods, and examples are illustrative only and are not intended to be limiting, unless otherwise specified.

[0096] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, and recombinant DNA, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (J. Abelson and M. Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols. 154 and 155 (Wu et al. eds.)

and Vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory).

Bacterium of the Invention

[0097] As indicated above, the present invention is based on the unexpected and surprising finding that methylo-trophic bacteria can efficiently produce lactate through the expression of a polypeptide having lactate dehydrogenase activity.

[0098] The present invention thus provides in a first aspect a genetically engineered methylo-trophic bacterium which has been modified to have an increased protein expression of a polypeptide having lactate dehydrogenase activity compared to an otherwise identical bacterium that does not carry said modification.

[0099] A "polypeptide having lactate dehydrogenase activity" is a polypeptide that catalyzes the reaction: lactate+NAD (+) \rightleftharpoons pyruvate+NADH. Non-limiting examples of such polypeptides are provided in SEQ ID Nos: 1 to 130 and 182 to 210. The polypeptide having lactate dehydrogenase activity may be a polypeptide having L-lactate dehydrogenase activity or a polypeptide having D-lactate dehydrogenase activity. A polypeptide having L-lactate dehydrogenase activity (EC 1.1.1.27) is a polypeptide that catalyzes the reaction: (S)-lactate+NAD (+) \rightleftharpoons pyruvate+NADH. Non-limiting examples of such polypeptides are provided in SEQ ID NO: 49 to 130. A polypeptide having D-lactate dehydrogenase activity (EC 1.1.1.28) is a polypeptide that catalyzes the reaction: (R)-lactate+NAD (+) \rightleftharpoons pyruvate+NADH. Non-limiting examples of such polypeptides are provided in SEQ ID NO: 1 to 48.

[0100] A "polypeptide having lactate dehydrogenase activity which is not allosterically regulated by fructose-1, 6-bisphosphate" is a polypeptide that catalyzes the LDH reaction above but which is not allosterically regulated by the metabolite fructose-1,6-bisphosphate. Non-limiting examples of such polypeptides are provided in SEQ ID NO: 182 to 210. Further, such polypeptides having lactate dehydrogenase activity which is not allosterically regulated by fructose-1,6-bisphosphate are also found as LDH polypeptides not having a histidine as the amino acid residue at the position corresponding to *L. casei* LDH Histidine 188, c.f. the procedure described in Example 8.

[0101] Polypeptides having lactate dehydrogenase activity are encoded in the genomes of a wide range of organisms. The polypeptide having lactate dehydrogenase may be derived from the same species as the bacterium in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to some embodiments, the polypeptide having lactate dehydrogenase activity is derived from the same species as the bacterium in which it is expressed. According to some embodiments, the polypeptide having lactate dehydrogenase activity is derived from a species different from the one in which it is expressed (i.e. it is heterologous).

[0102] The polypeptide having lactate dehydrogenase activity may be a functional variant of a naturally occurring polypeptide having lactate dehydrogenase activity, i.e. it may be a polypeptide having lactate dehydrogenase activity which differs from the naturally occurring polypeptide having lactate dehydrogenase activity in the amino acid composition. Such functional variant may comprise an amino

acid sequence which has at least 70%, such as at least 75%, at least 80%, at least 85%, at least 90% of at least 95%, sequence identity with the naturally occurring polypeptide and has lactate dehydrogenase activity.

[0103] According to some embodiments, the bacterium of the present invention expresses a heterologous polypeptide having lactate dehydrogenase activity.

[0104] By "increased protein expression" it is meant that the amount of the polypeptide having lactate dehydrogenase activity produced by the thus modified bacterium is increased compared to an otherwise identical bacterium that does not carry said modification. More particularly, by "increased expression" it is meant that the amount of the polypeptide having lactate dehydrogenase activity produced by the thus modified bacterium is increased by at least 10%, such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared to an otherwise identical bacterium that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry, or Western Blotting.

[0105] An increase in protein expression may be achieved by any suitable means well-known to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of a nucleotide sequence encoding the polypeptide having lactate dehydrogenase activity in the bacterium, such as by introducing into the bacterium at least one exogenous nucleic acid molecules comprising at least one nucleotide sequence encoding said polypeptide having lactate dehydrogenase activity.

[0106] Thus, according to some embodiments, the bacterium of the present invention comprises at least one exogenous nucleic acid molecule (such as a vector) comprising at least one nucleotide sequence encoding the polypeptide having lactate dehydrogenase activity.

[0107] Suitably, the exogenous nucleic acid molecule comprises at least one transcriptional unit comprising, from 5' to 3', a promoter that is functional in the bacterium to cause the production of an mRNA molecule and that is operably linked to a nucleotide sequence encoding said polypeptide having lactate dehydrogenase activity, and a transcriptional terminator sequence. The exogenous nucleic acid molecule may comprise at least two transcriptional units each comprising, from 5' to 3', a promoter that is functional in the bacterium to cause the production of an mRNA molecule and that is operably linked to a nucleotide sequence encoding said polypeptide, and a transcriptional terminator sequence. The transcriptional units may have the same type of promoter or different types of promoter.

[0108] The exogenous nucleic acid molecule may be a DNA construct, such as an expression cassette or a vector. The exogenous nucleic acid molecule may thus be a vector, such as an expression vector, or part of such vector, such as an expression cassette comprised by such vector. Normally, such a vector remains extrachromosomal within the bacterial cell which means that it is found outside of the genome of the bacterial cell. Alternatively, the exogenous nucleic acid

molecule may be stably integrated into the genome of the bacterium (e.g., by random or targeted insertion). Particularly, the exogenous nucleic acid molecule may be an expression cassette stably integrated into the genome of the bacterium (e.g., by random or targeted insertion).

[0109] An increase in protein expression may also be achieved by the integration of at least a second copy of the endogenous gene encoding the polypeptide having lactate dehydrogenase activity into the genome of the bacterium.

[0110] An increase in protein expression may also be achieved by increasing the strength of the promoter operably linked to the endogenous gene encoding the polypeptide having lactate dehydrogenase activity, e.g. by replacing the native promoter with a promoter that enables higher expression and overproduction of polypeptide compared to the native promoter. The promoters that can be used include natural promoters from *Bacillus subtilis*, *Bacillus amyloliquefaciens* or similar, such as P43, P15, Pveg, Pylb, PgroES, PsigX, PtrnQ, Ppst, PsodA, PrpsF, PlepA, PliaG, PrpsF, Ppst, PfusA, PsodA, Phag as well as artificial promoters active in *Bacillus subtilis* or inducible *Bacillus subtilis* promoters, such as PmtIA, Pspac, PxyIA, PsacB, or similar. Further examples include natural promoters from *Corynebacterium*, such as P CP_2454, Ptuf and Psod, natural promoters from *E. coli*, such as T7, ParaBAD, Plac, Ptac and Ptrc, and the promoter P F1 derived from the corynephage BFK20.

[0111] An increase in protein expression may also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the polypeptide having lactate dehydrogenase activity. By modifying the sequence of the ribosome binding site, the translation initiation rate may be increased, thus increasing translation efficiency.

[0112] A polypeptide having lactate dehydrogenase activity for use according to the invention may for instance be a polypeptide having lactate dehydrogenase activity selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 1 to 130; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 130.

[0113] According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises the amino acid sequence of any one of SEQ ID NOs: 1 to 130.

[0114] According to some embodiments, the "polypeptide having lactate dehydrogenase activity" is a polypeptide having L-lactate dehydrogenase activity.

[0115] A polypeptide having L-lactate dehydrogenase activity for use according to the invention may for instance be a polypeptide having lactate dehydrogenase activity selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 49 to 130; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 49 to 130.

[0116] According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 49 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 49 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 49 to 130. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises the amino acid sequence of any one of SEQ ID NOs: 49 to 130.

[0117] According to some embodiments, the "polypeptide having lactate dehydrogenase activity" is a polypeptide having D-lactate dehydrogenase activity.

[0118] A polypeptide having D-lactate dehydrogenase activity for use according to the invention may for instance be a polypeptide having lactate dehydrogenase activity selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 1 to 48; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 48.

[0119] According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 48. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 48.

[0120] According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence of any one of SEQ ID NOs: 1 to 48. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises the amino acid sequence of any one of SEQ ID NOs: 1 to 48.

[0121] According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises amino acid sequence having at least 70%, such as at least 75%, sequence identity with SEQ ID NO: 1. According to some embodi-

ments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with SEQ ID NO: 1. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with SEQ ID NO: 1. According to some embodiments, the polypeptide having lactate dehydrogenase activity comprises the amino acid sequence of SEQ ID NO: 1.

[0122] Techniques for determining lactate dehydrogenase activity are well known to the skilled person. The lactate dehydrogenase activity may for instance be determined in accordance with the following method:

[0123] The cells expressing lactate dehydrogenase are centrifuged and lysed to release the intracellular contents using B-PER (Thermo Scientific) according to manufacturer instructions. Briefly, 4 mL B-PER is mixed with 1 g wet cell biomass and incubated for 15 minutes at room temperature. After the cells have been disrupted, the lysate is clarified using centrifugation. The clear cell lysate is used for determination of lactate dehydrogenase activity. The enzymatic reaction is performed at 37° C. The reaction contains 2.8 mL of 0.13 mM (B NADH) prepared in 100 mM Sodium phosphate buffer (pH 7.5) and 0.1 mL of 34 mM sodium pyruvate prepared in 100 mM Sodium phosphate buffer (pH 7.5). The components are mixed, and A340 is measured until constant. After stabilization, 0.1 mL cell free extract is added to the reaction, and mixed. The decrease in A340 is continuously measured for 5 minutes. The linear slope of the reaction is used to calculate the lactate dehydrogenase activity according to the formula U/ml=(ΔA340/min)*3*dilution factor/(6.22*0.1).

[0124] In order to overcome certain drawbacks in bioprocesses, such as to avoid mass cell lysis, the bacterium of the present invention may be further modified to decrease the expression and/or activity of an endogenous polypeptide having polyphosphate kinase activity in bacteria.

[0125] Thus, according to some embodiments, a bacterium of the present invention has been further modified to have a decreased expression and/or activity of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

[0126] A "polypeptide having polyphosphate kinase activity" is a polypeptide that catalyzes the reaction: ATP+
(phosphate)_n↔ADP+(phosphate)_{n+1} (EC 2.7.4.1). Polyphosphate kinase (PPK), which is encoded by the ppk gene, is highly conserved in many bacteria, including methylo-trophs such as *Methylobacillus flagellatus* and *Methylobacillus glycogenes*, and plays a crucial role in the ability of bacteria to adapt to nutritional stringencies and environmental stresses. Non-limiting examples of an endogenous polypeptide having polyphosphate kinase activity are provided in SEQ ID NOs: 156, 158, 160, 162 and 164.

[0127] According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOs: 156, 158, 160, 162 and 164. According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOs: 156, 158, 160, 162 and

164. According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOS: 156, 158, 160, 162 and 164.

[0128] According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 70%, such as at least 75%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOS: 157, 159, 161, 163 and 165. According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 80%, such as at least 85%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOS: 157, 159, 161, 163 and 165. According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 90%, such as at least 95%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOS: 157, 159, 161, 163 and 165.

[0129] According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 156. According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 156. According to some embodiments, the endogenous polypeptide having polyphosphate kinase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 156.

[0130] According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 70%, such as at least 75%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 157. According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 80%, such as at least 85%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 157. According to some embodiments, the endogenous gene encoding the polypeptide having polyphosphate kinase activity comprises a nucleic acid sequence having at least 90%, such as at least 95%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 157.

[0131] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

[0132] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification. The expression level of the endogenous polypeptide having polyphosphate kinase activity may, for example, be decreased by at least 50%,

such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.

[0133] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression level of the endogenous gene encoding said endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification. The expression level of the endogenous gene may, for example, be decreased by at least 50%, such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.

[0134] According to some embodiments, the endogenous gene encoding said polypeptide having polyphosphate kinase activity has been inactivated, such as by deletion of part of or the entire gene sequence.

[0135] According to some embodiments, the endogenous gene encoding said polypeptide having polyphosphate kinase activity has been inactivated by introducing or expressing in the microorganism a rare-cutting endonuclease able to selectively inactivate by DNA cleavage, preferably by a double-strand break, the endogenous gene encoding said enzyme. A rare-cutting endonuclease to be used in accordance with the present invention to inactivate the endogenous gene may, for instance, be a transcription activator-like effector (TALE) nuclease, meganuclease, zinc-finger nuclease (ZFN), or RNA-guided endonuclease.

[0136] One way to inactivate the endogenous gene encoding said polypeptide having polyphosphate kinase activity is to use the CRISPRi system. The CRISPRi system was developed as a tool for targeted repression of gene expression or for blocking targeted locations on the genome. The CRISPRi system consists of the catalytically inactive, "dead" Cas9 protein (dCas9) and a guide RNA that defines the binding site for the dCas9 to DNA.

[0137] Thus, according to some embodiments, the endogenous gene encoding said polypeptide having polyphosphate kinase activity is inactivated by introducing or expressing in the bacterium an RNA-guided endonuclease, such as a catalytically inactive Cas9 protein, and a single guide RNA (sgRNA) specifically hybridizing (e.g. binding) under cellular conditions with the genomic DNA encoding a said polypeptide.

[0138] According to some embodiments, the expression of said endogenous polypeptide having polyphosphate kinase activity is decreased by way of inhibition.

[0139] Inhibition of the expression of the said endogenous polypeptide may be achieved by any suitable means known in the art. For example, the expression may be inhibited by gene silencing techniques involving the use of inhibitory nucleic acid molecules, such as antisense oligonucleotides, ribozymes, or interfering RNA (RNAi) molecules, such as microRNA (miRNA), small interfering RNA (siRNA), or short hairpin RNA (shRNA).

[0140] According to some embodiments, the expression of said endogenous polypeptide having polyphosphate kinase activity is decreased (e.g., inhibited) by transcriptional and/or translational repression of the endogenous gene encoding said polypeptide.

[0141] According to some embodiments, the expression of endogenous polypeptide having polyphosphate kinase activity is inhibited by introducing or expressing in the bacterium

an inhibitory nucleic acid molecule. For example, the inhibitory nucleic acid molecule may be introduced by way of an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said inhibitory nucleic acid molecule operably linked to a promoter, such as an inducible promoter, that is functional in the bacterium to cause the production of said inhibitory nucleic acid molecule. Suitably, the inhibitory nucleic acid molecule is one that specifically hybridizes (e.g. binds) under cellular conditions with cellular mRNA and/or genomic DNA encoding the endogenous polypeptide. Depending on the target, transcription of the encoding genomic DNA and/or translation of the encoding mRNA is/are inhibited.

[0142] According to some embodiments, the inhibitory nucleic acid molecule is an antisense oligonucleotide, ribozyme, or interfering RNA (RNAi) molecule. Preferably, such nucleic acid molecule comprises at least 10 consecutive nucleotides of the complement of the cellular mRNA and/or genomic DNA encoding the polypeptide or enzyme of interest (e.g., the cellular mRNA and/or genomic DNA encoding the polypeptide).

[0143] According to some embodiments, the inhibitory nucleic acid is an antisense oligonucleotide. Such antisense oligonucleotide is a nucleic acid molecule (either DNA or RNA), which specifically hybridizes (e.g. binds) under cellular conditions with the cellular mRNA and/or genomic DNA encoding the polypeptide.

[0144] According to some embodiments, the inhibitory nucleic acid molecule is a ribozyme, such as a hammerhead ribozyme. A ribozyme molecule is designed to catalytically cleave the mRNA transcript to prevent translation of the polypeptide.

[0145] According to some embodiments, the inhibitory nucleic acid molecule is an interfering RNA (RNAi) molecule. RNA interference is a biological process in which RNA molecules inhibit expression, typically destroying specific mRNA. Exemplary types of RNAi molecules include microRNA (miRNA), small interfering RNA (siRNA), and short hairpin RNA (shRNA). According to some embodiments, the RNAi molecule is a miRNA. According to some embodiments, the RNAi molecule is a siRNA. According to some embodiments, the RNAi molecule is an shRNA.

[0146] According to some embodiments, the bacterium of the invention has been modified to have a decreased activity of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

[0147] A decrease of the activity of the polypeptide having polyphosphate kinase activity may be achieved by any suitable means known in the art. For example, the activity may be decreased by introducing one or more mutations in the active site of the polypeptide resulting in the reduction or loss of activity. Thus, according to some embodiments, the activity of the endogenous polypeptide having polyphosphate kinase activity is decreased by at least one active-site mutation resulting in the reduction or loss of activity. At least one active-site mutation may, for example, be at least one non-conservative amino acid substitution.

[0148] By way of example, if the activity of the endogenous polypeptide having polyphosphate kinase activity is to be decreased in *Methylobacillus flagellatus*, the at least one active-site mutation may occur at any one of positions R379, S384, F492, P511, R568, R625, Q679, H439, and H458 in the amino acid sequence set forth in SEQ ID NO:

156 which form part of the active site. In case of orthologous polypeptides having polyphosphate kinase activity, the at least one active-site mutation may be at a position which corresponds to any one of positions R379, S384, F492, P511, R568, R625, Q678, H439, and H458 in the amino acid sequence set forth in SEQ ID NO: 156.

[0149] By way of another example, if the activity of the endogenous the polypeptide having polyphosphate kinase activity is to be decreased in *Methylobacillus glycogenes*, the at least one active-site mutation may occur at any one of positions R79, S84, F192, P211, R268, R325, Q378, H139, in the amino acid sequence set forth in SEQ ID NO: 158, which form part of the active site. In case of orthologous polypeptides having polyphosphate kinase activity, the at least one active-site mutation may be at a position which corresponds to any one of positions R79, S84, F192, P211, R268, R325, Q378, H139, H158 in the amino acid sequence set forth in SEQ ID NO: 158.

[0150] By way of another example, if the activity of the endogenous the polypeptide having polyphosphate kinase activity is to be decreased in *Methylobacillus rhizosphaerae*, the at least one active-site mutation may occur at any one of positions R379, S384, F492, P511, R568, R625, Q678, H439 and H458 in the amino acid sequence set forth in SEQ ID NO: 160, which form part of the active site. In case of orthologous polypeptides having polyphosphate kinase activity, the at least one active-site mutation may be at a position which corresponds to any one of positions R379, S384, F492, P511, R568, R625, Q678, H439 and H458 in the amino acid sequence set forth in SEQ ID NO: 160.

[0151] By way of another example, if the activity of the endogenous the polypeptide having polyphosphate kinase activity is to be decreased in *Methylobacterium organophilum*, the at least one active-site mutation may occur at any one of positions R392, S397, F505, P524, R581, R643, H452 and H471 in the amino acid sequence set forth in SEQ ID NO: 162, which form part of the active site. In case of orthologous polypeptides having polyphosphate kinase activity, the at least one active-site mutation may be at a position which corresponds to any one of positions R392, S397, F505, P524, R581, R643, H452 and H471 in the amino acid sequence set forth in SEQ ID NO: 162.

[0152] By way of another example, if the activity of the endogenous the polypeptide having polyphosphate kinase activity is to be decreased in *Methylorum extorquens*, the at least one active-site mutation may occur at any one of positions R451, S456, F564, P583, R640, R702, H511 and H530 in the amino acid sequence set forth in SEQ ID NO: 164, which form part of the active site. In case of orthologous polypeptides having polyphosphate kinase activity, the at least one active-site mutation may be at a position which corresponds to any one of positions R451, S456, F564, P583, R640, R702, H511 and H530 in the amino acid sequence set forth in SEQ ID NO: 164.

[0153] The resistance of such bacterium against cell lysis may be further improved, especially under certain conditions such as carbon limitation, by decreasing the expression and/or activity of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity.

[0154] Thus, according to some embodiments, the bacterium of the present invention may be further modified to have a decreased expression and/or activity of an endogenous polypeptide having Acyl-homoserine-lactone (AHL)

synthase activity compared to an otherwise identical bacterium that does not carry said modification.

[0155] A “polypeptide having Acyl-homoserine-lactone (AHL) synthase activity” is a polypeptide that catalyzes the reaction: An acyl-[acyl-carrier-protein]+S-adenosyl-L-methionine<=> [acyl-carrier-protein]+S-methyl-5'-thioadenosine+an N-acyl-L-homoserine lactone (EC 2.3.1.184). Acyl-homoserine lactones (AHLs) are small signaling molecules used by many Gram-negative bacteria for coordinating their behavior as a function of their population density. This process, based on the biosynthesis and the sensing of such molecular signals, and referred to as Quorum Sensing (QS), regulates various gene expressions, including growth, virulence, biofilms formation, and toxin production. Non-limiting examples of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity are provided in SEQ ID NOs: 168, 170, 172, 174, 176, 178 and 180.

[0156] According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOs: 168, 170, 172, 174, 176, 178 and 180. According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOs: 168, 170, 172, 174, 176, 178 and 180. According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence set forth in any one of SEQ ID NOs: 168, 170, 172, 174, 176, 178 and 180.

[0157] According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 70%, such as at least 75%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOs: 169, 171, 173, 175, 177, 179 and 181. According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 80%, such as at least 85%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOs: 169, 171, 173, 175, 177, 179 and 181. According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 90%, such as at least 95%, sequence identity with the nucleic acid sequence set forth in any one of SEQ ID NOs: 169, 171, 173, 175, 177, 179 and 181.

[0158] According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an amino acid sequence having at least 70%, such as at least 75%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 168. According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an amino acid sequence having at least 80%, such as at least 85%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 168. According to some embodiments, the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises an

amino acid sequence having at least 90%, such as at least 95%, sequence identity with the amino acid sequence set forth in SEQ ID NO: 168.

[0159] According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 70%, such as at least 75%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 169. According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 80%, such as at least 85%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 169. According to some embodiments, the endogenous gene encoding the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity comprises a nucleic acid sequence having at least 90%, such as at least 95%, sequence identity with the nucleic acid sequence set forth in SEQ ID NO: 169.

[0160] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

[0161] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification. The expression level of the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity may, for example, be decreased by at least 50%, such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.

[0162] According to some embodiments, the bacterium of the invention may be modified to have a decreased expression level of the endogenous gene encoding said endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification. The expression level of the endogenous gene may, for example, be decreased by at least 50%, such as by at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or at least 100% compared to the otherwise identical bacterium.

[0163] According to some embodiments, the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity has been inactivated, such as by deletion of part of or the entire gene sequence.

[0164] According to some embodiments, the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity has been inactivated by introducing or expressing in the microorganism a rare-cutting endonuclease able to selectively inactivate by DNA cleavage, preferably by a double-strand break, the endogenous gene encoding said enzyme. A rare-cutting endonuclease to be used in accordance with the present invention to inactivate the endogenous gene may, for instance, be a transcription activator-like effector (TALE) nuclease, meganuclease, zinc-finger nuclease (ZFN), or RNA-guided endonuclease.

[0165] One way to inactivate the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is to use the CRISPRi system. The CRISPRi system was developed as a tool for targeted repression of gene expression or for blocking targeted locations on the genome. The CRISPRi system consists of the catalytically inactive, “dead” Cas9 protein (dCas9) and a guide RNA that defines the binding site for the dCas9 to DNA.

[0166] Thus, according to some embodiments, the endogenous gene encoding said polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is inactivated by introducing or expressing in the bacterium an RNA-guided endonuclease, such as a catalytically inactive Cas9 protein, and a single guide RNA (sgRNA) specifically hybridizing (e.g. binding) under cellular conditions with the genomic DNA encoding a said polypeptide.

[0167] According to some embodiments, the expression of said endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased by way of inhibition.

[0168] Inhibition of the expression of the said endogenous polypeptide may be achieved by any suitable means known in the art. For example, the expression may be inhibited by gene silencing techniques involving the use of inhibitory nucleic acid molecules, such as antisense oligonucleotides, ribozymes, or interfering RNA (RNAi) molecules, such as microRNA (miRNA), small interfering RNA (siRNA), or short hairpin RNA (shRNA).

[0169] According to some embodiments, the expression of said endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased (e.g., inhibited) by transcriptional and/or translational repression of the endogenous gene encoding said polypeptide.

[0170] According to some embodiments, the expression of endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is inhibited by introducing or expressing in the bacterium an inhibitory nucleic acid molecule. For example, the inhibitory nucleic acid molecule may be introduced by way of an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said inhibitory nucleic acid molecule operably linked to a promoter, such as an inducible promoter, that is functional in the bacterium to cause the production of said inhibitory nucleic acid molecule. Suitably, the inhibitory nucleic acid molecule is one that specifically hybridizes (e.g. binds) under cellular conditions with cellular mRNA and/or genomic DNA encoding the endogenous polypeptide. Depending on the target, transcription of the encoding genomic DNA and/or translation of the encoding mRNA is/are inhibited.

[0171] According to some embodiments, the inhibitory nucleic acid molecule is an antisense oligonucleotide, ribozyme, or interfering RNA (RNAi) molecule. Preferably, such nucleic acid molecule comprises at least 10 consecutive nucleotides of the complement of the cellular mRNA and/or genomic DNA encoding the polypeptide or enzyme of interest (e.g., the cellular mRNA and/or genomic DNA encoding the polypeptide).

[0172] According to some embodiments, the inhibitory nucleic acid is an antisense oligonucleotide. Such antisense oligonucleotide is a nucleic acid molecule (either DNA or RNA), which specifically hybridizes (e.g. binds) under cellular conditions with the cellular mRNA and/or genomic DNA encoding the polypeptide.

[0173] According to some embodiments, the inhibitory nucleic acid molecule is a ribozyme, such as a hammerhead ribozyme. A ribozyme molecule is designed to catalytically cleave the mRNA transcript to prevent translation of the polypeptide.

[0174] According to some embodiments, the inhibitory nucleic acid molecule is an interfering RNA (RNAi) molecule. RNA interference is a biological process in which RNA molecules inhibit expression, typically destroying specific mRNA. Exemplary types of RNAi molecules include microRNA (miRNA), small interfering RNA (siRNA), and short hairpin RNA (shRNA). According to some embodiments, the RNAi molecule is a miRNA. According to some embodiments, the RNAi molecule is a siRNA. According to some embodiments, the RNAi molecule is an shRNA.

[0175] According to some embodiments, the bacterium of the invention has been modified to have a decreased activity of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

[0176] A decrease of the activity of the polypeptide having Acyl-homoserine-lactone (AHL) synthase activity may be achieved by any suitable means known in the art. For example, the activity may be decreased by introducing one or more mutations in the active site of the polypeptide resulting in the reduction or loss of activity. Thus, according to some embodiments, the activity of the endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity is decreased by at least one active-site mutation resulting in the reduction or loss of activity. At least one active-site mutation may, for example, be at least one non-conservative amino acid substitution.

[0177] As further detailed in Example 7, the present inventors have observed an unexpectedly high lactate titer when expressing a lactate dehydrogenase in a methylotrophic bacterium which utilizes the Eda-dependent RuMP pathway for methanol assimilation, such as *Methylobacillus flagellutes*. Accordingly, it will be advantageous to (over-)express a lactate dehydrogenase in a methylotrophic bacterium which expresses a RuMP cycle, and more specifically an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle.

[0178] Thus, according to some embodiments, the bacterium of the present invention expresses a RuMP cycle, preferably an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle.

[0179] A methylotrophic bacterium which expresses a RuMP cycle is a bacterium which expresses enzymes involved in the RuMP cycle. For example, a methylotrophic bacterium which expresses a RuMP cycle is a bacterium which expresses at least the following enzymes: a polypeptide having 3-hexulose-6-phosphate synthase activity (EC 4.1.2.43) and a polypeptide having 6-phospho-3-hexulose-isomerase activity (EC 5.3.1.27). The methylotrophic bacterium may inherently (i.e. natively) express a RuMP cycle or may be modified to express a RuMP cycle by using, e.g., DNA recombination techniques.

[0180] As used herein, a “polypeptide having 3-hexulose-6-phosphate synthase activity” means a polypeptide that catalyzes the reaction: D-arabino-hex-3-ulose 6-phosphate \rightleftharpoons D-ribulose 5-phosphate+formaldehyde (EC 4.1.2.43). A non-limiting example of such polypeptide is provided in SEQ ID NO: 135 and variants thereof having at

least 70%, such as at least 80%, at least 85%, at least 90%, at least 95% or at least 97%, sequence identity therewith.

[0181] As used herein, a “polypeptide having 6-phospho-3-hexuloseisomerase activity” means a polypeptide that catalyzes the reaction: D-arabino-hex-3-ulose 6-phosphate<=>D-fructose 6-phosphate (EC 5.3.1.27). A non-limiting example of such polypeptide is provided in SEQ ID NO: 136 and variants thereof having at least 70%, such as at least 80%, at least 85%, at least 90%, at least 95% or at least 97%, sequence identity therewith.

[0182] According to some embodiments, the bacterium of the present invention expresses a polypeptide having 3-hexulose-6-phosphate synthase activity and a polypeptide having 6-phospho-3-hexuloseisomerase activity.

[0183] According to some embodiments, the bacterium of the present invention expresses an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle.

[0184] A methylotrophic bacterium which expresses an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle is a bacterium which expresses, besides enzymes involved in the RuMP cycle, notably a polypeptide having 3-hexulose-6-phosphate synthase activity and a polypeptide having 6-phospho-3-hexuloseisomerase activity, also a polypeptide having phosphoglucoisomerase activity (EC 5.3.1.9) and a polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity (EC 4.1.2.14). The methylotrophic bacterium may inherently (i.e. natively) express a polypeptide having phosphoglucoisomerase activity and a polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity or may be modified to express said polypeptides by using, e.g., DNA recombination techniques.

[0185] As used herein, a “polypeptide having phosphoglucoisomerase activity” means a polypeptide that catalyzes the reaction: Alpha-D-glucose 6-phosphate<=>beta-D-fructofuranose 6-phosphate (EC 5.3.1.9). A non-limiting example of such polypeptide is provided in SEQ ID NO: 131 and variants thereof having at least 70%, such as at least 80%, at least 85%, at least 90%, at least 95% or at least 97%, sequence identity therewith.

[0186] As used herein, a “polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity” means a polypeptide that catalyzes the reaction: 2-dehydro-3-deoxy-6-phosphate-D-gluconate<=>pyruvate+D-glyceraldehyde 3-phosphate (EC 4.1.2.14). Non-limiting examples of such polypeptide are provided in SEQ ID NOs: 132 to 134 and variants thereof having at least 70%, such as at least 80%, at least 85%, at least 90%, at least 95% or at least 97%, sequence identity therewith.

[0187] According to some embodiments, the bacterium of the present invention (further) expresses a polypeptide having phosphoglucoisomerase activity and a polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity.

[0188] According to some embodiments, the bacterium of the present invention expresses a polypeptide having 3-hexulose-6-phosphate synthase activity, a polypeptide having 6-phospho-3-hexuloseisomerase activity, a polypeptide having phosphoglucoisomerase activity and a polypeptide having 2-keto-3-deoxy-phosphogluconate aldolase activity.

[0189] Non-limiting examples of methylotrophic bacteria which inherently (i.e. natively) express a RuMP cycle, and more specifically an Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle include *Methylobacillus* such as *Methylobacillus flagellatus* and *Methylobacillus glycogenes*.

[0190] Generally, a bacterium as referred to herein may be any suitable methylotrophic bacterium. The bacterium may be Gram-positive or Gram-negative. Preferably, the bacterium is a Gram-negative bacterium.

[0191] According to some embodiments, the bacterium of the present invention is a mesophilic, methylotrophic bacterium.

[0192] According to some embodiments, the bacterium of the present invention belongs to the family Methylophilaceae or Methylobacteriaceae.

[0193] According to some embodiments, the bacterium of the present invention belongs to the family Methylophilaceae.

[0194] According to some embodiments, the bacterium of the present invention belongs to the family Methylobacteriaceae.

[0195] According to some embodiments, the bacterium of the present invention belongs to the genus *Methylobacillus*, *Methylobacterium* or *Methylorubrum*.

[0196] According to some embodiments, the bacterium of the present invention belongs to the genus *Methylobacillus* or *Methylobacterium*.

[0197] According to some embodiments, the bacterium of the present invention belongs to the genus *Methylobacterium* or *Methylorubrum*.

[0198] According to some embodiments, the bacterium of the present invention is selected from *Methylobacillus flagellatus*, *Methylobacillus glycogenes*, *Methylobacillus pratensis*, *Methylobacillus rhizosphaerae*, *Methylobacillus gramineus*, *Methylobacillus arboreus*, *Methylobacillus caricae*, *Methylobacillus methilovorans*, *Methylobacillus* sp., *Methylobacterium extorquens*, *Methylobacterium organophilum* and *Methylorubrum extorquens*.

[0199] According to some embodiments, the bacterium of the present invention belongs to the genus *Methylobacillus*.

[0200] According to some embodiments, the bacterium of the present invention is selected from *Methylobacillus flagellatus*, *Methylobacillus glycogenes*, *Methylobacillus pratensis*, *Methylobacillus rhizosphaerae*, *Methylobacillus gramineus*, *Methylobacillus arboreus*, *Methylobacillus caricae*, *Methylobacillus methilovorans*, and *Methylobacillus* sp.

[0201] According to some embodiments, the bacterium of the present invention is *Methylobacillus flagellutes*.

[0202] According to some embodiments, the bacterium of the present invention is *Methylobacillus glycogenes*.

[0203] According to some embodiments, the bacterium of the present invention is *Methylobacillus rhizosphaerae*.

Method of the Invention

[0204] The present invention also provides methods for producing lactate comprising cultivating a bacterium according to the invention under suitable culture conditions. The method may further comprise collecting lactate from the culture medium.

[0205] According to some embodiments, the present invention provides a method for producing L-lactate.

[0206] According to some embodiments, the present invention provides a method for producing D-lactate.

[0207] According to some embodiments, the present invention provides a method for producing L-lactate and D-lactate.

[0208] The culture medium employed may be any conventional medium suitable for culturing a bacterium cell in

question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective bacterium, such as carbon and nitrogen sources and other inorganic salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as *E. coli* cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others.

[0209] The carbon source may be any suitable carbon substrate known in the art, and in particularly any carbon substrate commonly used in the cultivation of methyotrophic bacteria and/or fermentation. A carbon source of particular interest is a reduced one-carbon compound, such as methanol, methane, formate, or methylamine, or a multi-carbon compound that contains no carbon-carbon bonds, such as dimethyl ether and dimethylamine. Thus, according to some embodiments, the culture medium comprises methanol as a carbon source. The concentration of methanol in the culture medium may generally be between in the range from about 0.5% w/v to about 4% w/v, such as from about 2% w/v to about 4% w/v. According to some embodiments, the concentration of methanol in the culture medium is in the range from about 2.5% w/v to about 3.5% w/v.

[0210] As the nitrogen source, various ammonium salts such as ammonia and ammonium sulfate, other nitrogen compounds such as amines, a natural nitrogen source such as peptone, soybean-hydrolysate, and digested fermentative microorganism can be used. As minerals, potassium monophosphate, magnesium sulfate, sodium chloride, ferrous sulfate, manganese sulfate, calcium chloride, and the like can be used.

[0211] Suitably, the bacterium is cultivated under suitable conditions for the production of the desired product. Suitable conditions for culturing the respective bacterium are well known to the skilled person. Typically, a bacterium is cultured at a temperature ranging from about 20 to about 45° C., such as from about 30 to about 38° C., such as at about 37° C. The cultivation can be preferably performed under aerobic conditions, such as by a shaking culture, by a stirring culture or in a bioreactor with aeration, at a temperature of about 20 to about 45° C., such as about 30 to 38° C., preferably at about 37° C. The pH of the culture is usually above 5, such as in a range from about 6 to about 8, preferably from about 6.5 to about 7.5, more preferably from about 6.8 to about 7.2.

[0212] The pH of the culture can be adjusted with ammonia, calcium carbonate, various acids, various bases, and buffers. The cultivation may be carried out for a period in the range from 10 to 70 h, preferably in a range from 24 to 60 h, more preferably in a range from 36 to 50 h.

[0213] After cultivation, solids such as cells can be removed from the culture medium by centrifugation or membrane filtration. The biochemical compound can be collected by conventional method for isolation and purification chemical compounds from a medium. Well-known

purification procedures include, but are not limited to, centrifugation or filtration, precipitation, ion exchange, chromatographic methods such as e.g. ion exchange chromatography or gel filtration chromatography, and crystallization methods. The method may further comprise collecting lactate from the culture medium.

[0214] The present invention also provides a biochemical compound obtainable by a method as detailed above.

Certain Other Definitions

[0215] The term "mesophilic" as used herein in the context of a bacterium means that the bacterium grows best in moderate temperature with an optimum growth range from 20 to 45° C.

[0216] The term "methyotrophic" as used herein in the context of a bacterium means that the bacterium can use reduced one-carbon compounds, such as methanol, methane, formate, or methylamine, as the carbon source for their growth, and multi-carbon compounds that contain no carbon-carbon bonds, such as dimethyl ether and dimethylamine.

[0217] "Polypeptide" and "protein" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristylation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

[0218] "Nucleic acid" or "polynucleotide" are used interchangeably herein to denote a polymer of at least two nucleic acid monomer units or bases (e.g., adenine, cytosine, guanine, thymine) covalently linked by a phosphodiester bond, regardless of length or base modification.

[0219] "Recombinant" or "non-naturally occurring" when used with reference to, e.g., a host cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant bacterial cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

[0220] "Heterologous" or "exogenous" as used herein in the context of a gene or nucleic acid molecule refer to a gene or nucleic acid molecule (i.e. DNA or RNA molecule) that does not occur naturally as part of the genome of the bacterium in which it is present or which is found in a location or locations in the genome that differ from that in which it occurs in nature. Thus, a "heterologous" or "exogenous" gene or nucleic acid molecule is not endogenous to the bacterium and has been exogenously introduced into the microorganism. A "heterologous" gene or nucleic acid molecule DNA molecule may be from a different organism, a different species, a different genus or a different kingdom, as the host DNA.

[0221] "Heterologous" as used herein in the context of a polypeptide means that a polypeptide is normally not found in or made (i.e. expressed) by the host microorganism, but derived from a different organism, a different species, a different genus or a different kingdom.

[0222] As used herein, the term “ortholog” or “orthologs” refers to genes, nucleic acid molecules encoded thereby, i.e., mRNA, or proteins encoded thereby that are derived from a common ancestor gene but are present in different species.

[0223] By “decreased expression” of a gene it is meant that the amount of the transcription product, respectively the amount of the polypeptide (e.g., enzyme) encoded by said gene produced by the modified bacterium is decreased compared to an otherwise identical bacterium that does not carry said modification. More particularly, by “decreased expression” of a gene it is meant that the amount of the transcription product, respectively the amount of the polypeptide (e.g., enzyme) encoded by said gene produced by the modified bacterium is decreased by at least 10%, such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90% or at least 100%, compared to an otherwise identical bacterium that does not carry said modification. The level of expression of a gene can be determined by well-known methods, including PCR, Southern blotting, and the like. In addition, the level of gene expression can be estimated by measuring the amount of mRNA transcribed from the gene using various well-known methods, including Northern blotting, quantitative RT-PCR, and the like. The amount of the polypeptide encoded by the gene can be measured by well-known methods, including ELISA, Immunohistochemistry or Western Blotting and the like.

[0224] Expression of a gene can be decreased by introducing a mutation into the gene in the genome of the bacterium so that the intracellular activity of the polypeptide encoded by the gene is decreased as compared to an otherwise identical bacterium that does not carry said mutation. Mutations which result in a decreased expression of the gene include the replacement of one nucleotide or more to cause an amino acid substitution in the polypeptide encoded by the gene (missense mutation), introduction of a stop codon (nonsense mutation), deletion or insertion of nucleotides to cause a frame shift, insertion of a drug-resistance gene, or deletion of a part of the gene or the entire gene (Qiu and Goodman, 1997; Kwon et al., 2000). Expression can also be decreased by modifying an expression regulating sequence such as the promoter, the Shine-Dalgarno (SD) sequence, etc. Expression of the gene can also be decreased by gene replacement (Datsenko and Wanner, 2000), such as the “lambda-red mediated gene replacement”. The lambda-red mediated gene replacement is a particularly suitable method to inactive one or more genes as described herein.

[0225] “Inactivating”, “inactivation” and “inactivated”, when used in the context of a gene or gene cluster, means that the gene or gene cluster in question no longer expresses a functional protein. It is possible that the modified DNA region is unable to naturally express the gene or gene cluster due to the deletion of a part of or the entire sequence of the gene or gene cluster, the shifting of the reading frame of the gene or gene cluster, the introduction of missense/nonsense mutation(s), or the modification of the regulatory region of the gene or gene cluster, including sequences controlling gene expression, such as a promoter, enhancer, attenuator, ribosome-binding site, etc. Preferably, a gene or gene cluster of interest is inactivated by deletion of a part of or the entire sequence of the gene or gene cluster, such as by gene replacement. Inactivation may also be accomplished by introducing or expressing a rare-cutting endonuclease able to selectively inactivating by DNA cleavage, preferably by

double-strand break, the gene or gene cluster of interest. A “rare-cutting endonuclease” within the context of the present invention includes transcription activator-like effector (TALE) nucleases, meganucleases, zing-finger nucleases (ZFN), and RNA-guided endonucleases.

[0226] The presence or absence of a gene or gene cluster in the genome of a bacterium can be detected by well-known methods, including PCR, Southern blotting, and the like. In addition, the level of gene expression can be estimated by measuring the amount of mRNA transcribed from the gene or gene cluster using various well-known methods, including Northern blotting, quantitative RT-PCR, and the like. The amount of the polypeptide encoded by the gene or gene cluster can be measured by well-known methods, including SDS-PAGE followed by an immunoblotting assay (Western blotting analysis), and the like.

[0227] As used herein, “decreased”, “decreasing” or “decrease of” expression of a polypeptide (such as a polypeptide as described herein) means that the expression of said polypeptide in a modified bacterium is reduced compared to the expression of said polypeptide in an otherwise identical bacterium that does not carry said modification (control). The expression of a polypeptide in a modified bacterium may be reduced by at least about 10%, and preferably by at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 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%, at least about 99% or 100%, or any percentage, in whole integers between 10% and 100% (e.g., 6%, 7%, 8%, etc.), compared to the expression of said polypeptide in an otherwise identical bacterium that does not carry said modification (control). More particularly, “decreased”, “decreasing” or “decrease of” expression of a polypeptide means that the amount of the polypeptide in the modified bacterium is reduced by at least about 10%, and preferably by at least about 20%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, 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%, at least about 99% or 100%, or any percentage, in whole integers between 10% and 100% (e.g., 6%, 7%, 8%, etc.), compared to the amount of said polypeptide in an otherwise identical bacterium that does not carry said modification (control). The expression or amount of a polypeptide in a bacterium can be determined by any suitable means known in the art, including techniques such as ELISA, Immunohistochemistry, Western Blotting or Flow Cytometry.

[0228] As used herein, “abolished” expression of a polypeptide (such as a polypeptide as described herein) means that the expression of said polypeptide in a modified bacterium is not detectable compared to the expression of said polypeptide in an otherwise identical bacterium that does not carry said modification (control).

[0229] As used herein, “decreased”, “decreasing” or “decrease of” activity of a polypeptide (such as an enzyme as described herein) means that the catalytic activity of said polypeptide in a modified bacterium is reduced compared to the catalytic activity of said polypeptide in an otherwise identical bacterium that does not carry said modification (control). The activity of a polypeptide in a modified bac-

terium may be reduced by at least about 10%, and preferably by at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 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%, at least about 99% or 100%, or any percentage, in whole integers between 10% and 100% (e.g., 6%, 7%, 8%, etc.), compared to the expression of said polypeptide in an otherwise identical bacterium that does not carry said modification (control). The activity of a polypeptide in a bacterium can be determined by any suitable protein and enzyme activity assay.

[0230] “Expression” includes any step involved in the production of a polypeptide (e.g., encoded enzyme) including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

[0231] As used herein, “regulatory region” of a gene or gene cluster refers to a nucleic acid sequence that affect the expression of a coding sequence. Regulatory regions are known in the art and include, but are not limited to, promoters, enhancers, transcription terminators, polyadenylation sites, matrix attachment regions and/or other elements that regulate expression of a coding sequence.

[0232] “Substitution” or “substituted” refers to modification of the polypeptide by replacing one amino acid residue with another, for instance the replacement of an Serine residue with a Glycine or Alanine residue in a polypeptide sequence is an amino acid substitution. When used with reference to a polynucleotide, “substitution” or “substituted” refers to modification of the polynucleotide by replacing one nucleotide with another, for instance the replacement of a cytosine with a thymine in a polynucleotide sequence is a nucleotide substitution.

[0233] “Conservative substitution”, when used with reference to a polypeptide, refers to a substitution of an amino acid residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid, e.g., alanine, valine, leucine, and isoleucine; an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain, e.g., serine and threonine; an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain, e.g., phenylalanine, tyrosine, tryptophan, and histidine; an amino acid with a basic side chain is substituted with another amino acid with a basic side chain, e.g., lysine and arginine; an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain, e.g., aspartic acid or glutamic acid; and a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

[0234] “Non-conservative substitution”, when used with reference to a polypeptide, refers to a substitution of an amino acid in a polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., serine for glycine), (b) the charge or hydrophobicity, or (c) the bulk

of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0235] As used herein, “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double stranded nucleic acid loop into which additional nucleic acid segments can be ligated. Certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “expression vectors”. Certain other vectors are capable of facilitating the insertion of an exogenous nucleic acid molecule into a genome of a bacterium. Such vectors are referred to herein as “transformation vectors”. In general, vectors of utility in recombinant nucleic acid techniques are often in the form of plasmids.

[0236] In the present specification, “plasmid” and “vector” can be used interchangeably as the plasmid is the most commonly used form of a vector. Large numbers of suitable vectors are known to those of skill in the art and commercially available.

[0237] As used herein, “promoter” refers to a sequence of DNA, usually upstream (5') of the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription. The selection of the promoter will depend upon the nucleic acid sequence of interest. A suitable “promoter” is generally one which is capable of supporting the initiation of transcription in a bacterium of the invention, causing the production of an mRNA molecule.

[0238] As used herein, “operably linked” refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence “operably linked” to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence. A promoter sequence is “operably-linked” to a gene when it is in sufficient proximity to the transcription start site of a gene to regulate transcription of the gene.

[0239] “Percentage of sequence identity,” “% sequence identity” and “percent identity” refers to sequence identity between a nucleotide sequence and a reference nucleotide sequence or between an amino acid sequence and a reference amino acid sequence. Sequence identity can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are identical at that position. A degree of identity between nucleotide or amino acid sequences is a function of the number of identical or matching nucleotides or amino acids at positions shared by the nucleotide or amino acid sequences, respectively. Various alignment algorithms and/or programs may be used to calculate the identity between two sequences, including FASTA or BLAST which are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with default settings.

[0240] “Reference sequence” or “reference amino acid sequence” refers to a defined sequence to which another

sequence is compared. In the context of the present invention a reference amino acid sequence may, for example, be an amino acid sequence set forth in SEQ ID NO: 1.

[0241] As used herein, the term “about” means plus or minus 10% of the numerical value of the number with which it is being used.

[0242] Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and sub ranges within a numerical limit or range are specifically included as if explicitly written out.

[0243] As used herein, the indefinite articles “a” and “an” mean “at least one” or “one or more” unless the context clearly dictates otherwise.

[0244] As used herein, the terms “comprising”, “including”, “having” and grammatical variants thereof are to be taken as specifying the stated features, steps or components but do not preclude the addition of one or more additional features, steps, components or groups thereof.

[0245] Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

Example 1: Bioinformatic Analysis

[0246] While investigating the potential of *Methylobacillus flagellatus* for use in industrial biotechnology, we noticed it produced significantly higher titers of lactic acid under certain bioreactor fermentation conditions compared to other common overflow products such as acetate. This was unexpected because the published *Methylobacillus flagellatus* genome (Chistoserdova L. et al., 2007) contained no annotation for a lactate dehydrogenase enzyme. To identify which gene was responsible for lactic acid production the amino acid sequences of two well-characterized lactate dehydrogenase genes from *Escherichia coli*, IdhA (SEQ ID NO: 9) and IldD (SEQ ID NO: 49), were used to perform a BLAST search against the *Methylobacillus flagellatus* genome. The search resulted in only three hits, summarized in Table 1. The hit described as “2-hydroxyacid dehydrogenase” with the locus tag Mfla_0399 (SEQ ID NO: 137, encoding SEQ ID NO: 1) was deemed as the only plausible candidate.

TABLE 1

Summary of BLAST search of *E. coli* lactate dehydrogenases against the *Methylobacillus flagellatus* genome.

Query	Hit description	Query Coverage	Identity
IdhA	2-hydroxyacid dehydrogenase	100%	51%
IdhA	phosphoglycerate dehydrogenase	70%	30%
IldD	FMN-binding glutamate synthase family protein	28%	34%

[0247] The amino acid sequence of this gene was used to perform a BLAST search against all known genomes in the *Methylobacillus* genus, resulting in a similar “2-hydroxyacid dehydrogenase” hit in each genome (SEQ ID NOS: 2-8) with a high degree of similarity. A protein alignment using the Clustal Omega service was performed on the identified sequenced, revealing a high degree of conservation across

all strains, with several 100% conserved regions and an overall similarity of over 84%.

Example 2: Expression of Novel Identified Enzyme and Production of Lactate in Methylotrophs

[0248] To test whether the identified gene (Mfla_0399, SEQ ID NO: 137) is responsible for lactate production by *Methylobacillus flagellatus*, we prepared constructs that enable controlled gene expression in *M. flagellatus* and other *Methylobacilli*. The gene was amplified from the genome of *M. flagellatus* by PCR and cloned into an expression vector under the control of the IPTG-inducible lacO/trc promotor, using the NEB Hi-Fi cloning kit according to manufacturer instructions. The prepared plasmids were checked by sequencing and transformed into *M. flagellatus* via electroporation to create strain OCB 354.

[0249] To test lactate production, the transformants were cultivated in 250 ml shake flasks in a mineral medium mineral medium containing methanol, KH₂PO₄, Na₂HPO₄, MgSO₄, NH₄SO₄, and trace elements. The shake flasks were inoculated with 5 mL liquid overnight culture. After 4 h of growth, 0.5 g/L IPTG was added to induce gene expression. The cultures were sampled after 24 h of growth. Lactate was measured by HPLC.

[0250] The strain OCB 354, expressing the Mfla_0399 gene (SEQ ID NO: 137) produced 2 g/L lactate from methanol after 24 h of cultivation (FIG. 1). Under the same conditions, the control strain that did not express Mfla_0399 did not produce detectable amounts of lactic acid. This confirmed that Mfla_0399 is responsible for converting methanol into lactate in *M. flagellatus*.

Example 3: Expression of L-Lactate Dehydrogenase and Production of L-Lactate from Methanol

[0251] To enable L-lactate synthesis from methanol, we expressed a L-lactate dehydrogenase and expressed it in *M. flagellatus*. The construct and strains were prepared as the described in Example 2. The gene sequence was taken *Pediococcus* (strain OCB 457, SEQ ID NO: 50) and harmonized for gene expression in *M. flagellatus*.

[0252] To test lactate production, the transformants were cultivated in 250 ml shake flasks in a mineral medium mineral medium containing methanol, KH₂PO₄, Na₂HPO₄, MgSO₄, NH₄SO₄, and trace elements. The shake flasks were inoculated with 5 mL liquid overnight culture. After 4 h of growth, 0.5 g/L IPTG was added to induce gene expression. The cultures were sampled after 24 h of growth. Lactate was measured by HPLC.

[0253] The strain OCB 457 produced approximately 1.6 g/L lactate from methanol after 24 h of cultivation (FIG. 2). Under the same conditions, the control strain that did not express any lactate dehydrogenase gene did not produce detectable amounts of lactic acid.

Example 4: Determination of Lactic Acid Chirality

[0254] We tested the chirality of the lactate produced by strains OCB 354 and OCB 457. We tested the clarified cell broth at the end of the fermentation by using the enzymatic L-lactic acid and D-lactic acid specific kits from Megazyme (K-LATE and K-DATE, respectively). The assays were performed according to the manufacturer instructions. We confirmed that the chirality of the lactate was determined by

the enzyme. One racemate (either D-lactate or L-lactate) constituted more than 97% of the lactate, depending on the LDH enzyme (Table 2).

TABLE 2

Chirality determination of the lactic acid produced by various strains			
Strain	Enzyme	D-lactate	L-lactate
OCB 354	Mfla_0399 D-LDH	98.3%	1.7%
OCB 457	Pediococcus L-LDH	3.0%	97.0%

Example 5: Identification and Selection of Other Lactate Dehydrogenase Enzymes

[0255] Lactate dehydrogenases are very common and well-characterized enzymes present in a broad variety of organisms, from archaea to humans. While the conversion of pyruvate always requires reducing power, it can be provided in the form of different reducing equivalent cofactors. NADH dependent lactate dehydrogenases are the most biotechnologically relevant. To identify examples of lactate dehydrogenases covering a broad phylogenetic range, the E.C. numbers 1.1.1.27 (L-lactate dehydrogenases) and 1.1.1.28 (D-lactate dehydrogenases) were used to perform a search in the UniProt public protein database, generating 16 836 and 3 719 hits respectively. The complete results were downloaded and processed to group them by phylum of the source organism. The sequences were filtered for outliers regarding their length and a single example from each phylum was extracted, for a total of 41 D-lactate specific (SEQ ID NOs: 12-48, protein only) and 84 L-lactate specific (SEQ ID NOs: 51-130, protein only) NADH lactate dehydrogenase sequences.

Example 7: Metabolic Probing

[0256] When we expressed the lactate dehydrogenase, we observed an unexpectedly high lactate titer in the strain OCB 354. We therefore probed the central carbon metabolism of the organism by deleting several genes of the glycolytic and RuMP pathway.

[0257] While we were successful in deleting a variety of the genes, we could not delete the phosphoglucoisomerase (pgi) gene (Mfla_1325, gene sequence SEQ ID NO: 150, protein sequence SEQ ID NO: 131) in the strain. We investigated the genome of the organism for more details and found that this is critical for the Eda (2-keto-3-deoxy-phosphogluconate aldolase)-dependent RuMP cycle used by *M. flagellates* to extract C3 compounds from the RuMP cycle. This enzyme (Mfla_0760, protein sequence SEQ ID NO: 132, gene sequence SEQ ID NO: 151) splits the C6 compound 2-keto-3-deoxy-phosphogluconate (2,3-KDPG) into glyceraldehyde-3P (GAP) and pyruvate. GAP is used to replenish the RuMP cycle for the next cycle of methanol assimilation, while pyruvate exits the cycle (FIG. 3). Coincidentally, pyruvate is the metabolite needed for lactate formation by lactate dehydrogenase. This means that the strain utilizing the Eda-RuMP pathway for methanol assimilation has a surprisingly high, and direct metabolic flux towards pyruvate, which results in high lactate production observed in the strain OCB 354.

[0258] The alternative RuMP cycle depends on the fructose bisphosphate aldolase (Fba) to cleave the C6 interme-

diates of the RuMP cycle. This generates two GAP moieties, one of which is used for RuMP regeneration, and the other exists the cycle (FIG. 4). The GAP is then converted to pyruvate via a series of metabolic reactions. This could potentially be used for lactate production. However, most of the pathway intermediates can be used for biomass and byproduct formation, making the Fba-RuMP cycle less efficient for lactate formation from methanol, and the Eda-RuMP superior in comparison.

[0259] The combination of the Eda-dependent RuMP cycle and over-expression of lactate dehydrogenase is therefore beneficial for efficient lactate formation in strain OCB 354. A similar effect of increased lactate production is also expected when the lactate dehydrogenase (over-)expression is combined with the expression of heterologous Eda enzymes, for example from *Pseudomonas putida* or *E. coli* (SEQ ID NOs: 133 and 134, respectively).

Example 8: Bioinformatic Analysis of L-LDH

[0260] It has been reported that most L-lactate dehydrogenase enzymes are allosterically regulated by the metabolite fructose-1,6-bisphosphate (1,6-FBP), a key metabolite in glycolysis. However, since lactic acid synthesis and most of the core metabolism in *Methylobacillus flagellatus* is routed through the EDA RuMP pathway, which does not include 1,6-FBP, it is reasonable to assume this metabolite is not present in sufficient concentration to allosterically activate L-LDH. It was found that a key residue Histidine 188 controls this activation mechanism in *Lactobacillus casei* L-LDH (DOI: 10.1271/bbb.59.451). To identify additional non-allosteric L-LDH enzymes, the UniProt database was queried for all enzymes predicted to have L-LDH function (EC 1.1.1.27). The query returned roughly 12 000 amino acid sequences. The sequences were aligned using the MAFT protein alignment algorithm. The aligned sequences were filtered based on the amino acid at position corresponding to *L. casei* LDH Histidine 188 to exclude any sequence with a His at this position. This resulted in a shorter list of roughly 3700 sequences. To cover a wide range of sequences, a similarity tree (FIG. 5) of these sequences was constructed and visualized. 28 individual groups were manually identified from the tree and a single sequence from each group was randomly selected for screening, resulting in sequences SEQ ID 182-210. These sequences were optimized for expression in *Methylobacillus flagellatus* using a proprietary codon optimization algorithm written in the Python programming language, synthesized and expressed in *Methylobacillus flagellatus* as described in examples 3 and 4. To test lactate production, the transformants were cultivated in 250 ml shake flasks in a mineral medium containing methanol, KH₂PO₄, Na₂HPO₄, MgSO₄, NH₄SO₄, and trace elements. The shake flasks were inoculated with 5 mL liquid overnight culture. After 4 h of growth, 0.5 g/L IPTG was added to induce gene expression. The cultures were sampled after 24 h of growth. Lactate was measured by HPLC. Results are summarized in FIG. 6. Strains and sequences are summarized in Table 3.

TABLE 3

L-LDH sequences, shortnames and strains expressing same.		
SEQ ID	ShortName	strain
182	Akr	OCB 675
183	Lsp	OCB 676
184	Tsp	OCB 677
185	Wvi	OCB 678
186	Ror	OCB 679

TABLE 3-continued

L-LDH sequences, shortnames and strains expressing same.		
SEQ ID	ShortName	strain
187	Cje	OCB 680
188	Fba	OCB 681
189	Sal	OCB 682
190	Vvu	OCB 683
191	Ssp	OCB 684
192	Cae	OCB 685
193	Lam	OCB 686
194	Bav	OCB 687
195	Eti	OCB 688
196	Lba	OCB 689
197	Lka	OCB 690
198	Lri	OCB 691
199	Wci	OCB 693
200	Lhe	OCB 694
201	Cin	OCB 695
202	Sau	OCB 696
203	Lme	OCB 697
204	Lfe	OCB 698
205	Pod	OCB 699
206	Sau2	OCB 700
207	Flw	OCB 701
208	Vdo	OCB 702
209	Lhi	OCB 703
210	Lpl	OCB 704

Example 9: Production of L-Lactic Acid in Bioreactor

[0261] Cultures of *Methylobacillus flagellatus* expressing different lactate dehydrogenase enzymes were cultured in 5 liter bioreactors. A mineral medium containing KH_2PO_4 , Na_2HPO_4 , MgSO_4 , NH_4SO_4 and a trace element mixture was used. Methanol concentration was dynamically maintained at 4-5 g/L throughout the fermentations via a feedback loop using an on-line proprietary methanol sensor. The

pH was kept at 7 by automatic addition of ammonium hydroxide, which also served as source of nitrogen. Samples from both bioreactors were taken in regular intervals and analysed for lactate content by HPLC. Surprisingly, strain OCB 457 expressing the *Pediococcus* L-LDH, which produced 1.6 g/L of lactic acid in shaker-scale experiments (example 4) only produced 2.6 g/L of lactic acid within 48 hours of fermentation. One of the strains described in Example 8 was tested a 5 liter bioreactor using the same process protocol. Surprisingly, this strain produced ~25 g/L of lactic acid within 48 hours, over 10x more than strain Ocb 457, even though OCB 457 produced 2x more lactate in the shaker (FIG. 7).

LIST OF REFERENCES CITED IN THE DESCRIPTION

- [0262] Chistoserdova, L. et al. (2007) 'Genome of *Methylobacillus flagellatus*, Molecular Basis for Obligate Methylotrophy, and Polyphyletic Origin of Methylotrophy', *Journal of Bacteriology*. American Society for Microbiology Journals, 189 (11): 4020-4027.
- [0263] Qiu Z and Goodman M F: The *Escherichia coli* polB locus is identical to dinA, the structural gene for DNA polymerase II. Characterization of Pol II purified from a polB mutant. *J Biol Chem.* 1997, 272 (13): 8611-8617.
- [0264] Kwon D H, Peña J A, Osato M S, Fox J G, Graham D Y, Versalovic J: Frameshift mutations in rdxA and metronidazole resistance in North American *Helicobacter pylori* isolates. *J Antimicrob Chemother* 2000, 46 (5): 793-796
- [0265] Datsenko K A, Wanner B L: One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 2000, 97:6640-6645.

SEQUENCE LISTING

```
Sequence total quantity: 210
SEQ ID NO: 1      moltype = AA  length = 333
FEATURE          Location/Qualifiers
source           1..333
mol_type = protein
organism = Methylobacillus flagellatus

SEQUENCE: 1
MRIVFFSTQV YDRESFLALP RLQNAEFIGQ QPKLTLDAV LATGAEVVCA FINDDLSAPV 60
LERLAAANGTK LIALRSAGYN HIDLEAAQRL NLAVVVRVPAY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVHKTVGII GTGQIGAAFA RIMAGFGCRI LAYDPYPNPE 180
VIALGASYLE LDTLLAQSHI VSLHCPLSPA THHLINAGSL SRMQPGSMLI NTSRGALVDT 240
PAVIEALKSG HLGYLGLDVY EEEADLFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALNIAST TLDNISAWAA GHPQNLVRLE TGS 333

SEQ ID NO: 2      moltype = AA  length = 333
FEATURE          Location/Qualifiers
source           1..333
mol_type = protein
organism = Methylobacillus sp

SEQUENCE: 2
MRIVFFSTQV YDRESFLALP RLQNAEFIGQ QPKLTLDAV LATGAEVVCA FINDDLSAPV 60
LERLAAANGTK LIALRSAGYN HIDLEAAQRL NLAVVVRVPAY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVHKTVGII GTGQIGAAFA RIMAGFGCRI LAYDPYPNPE 180
VIALGASYLE LDTLLAQSHI VSLHCPLSPA THHLINAGSL SRMQPGSMLI NTSRGALVDT 240
PAVIEALKSG HLGYLGLDVY EEEADLFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALNIAST TLDNISAWAA GHPQNLVRLE TGS 333

SEQ ID NO: 3      moltype = AA  length = 331
FEATURE          Location/Qualifiers
```

-continued

```

source          1..331
               mol_type = protein
               organism = Methylobacillus rhizosphaerae

SEQUENCE: 3
MRIVFFSTQV YDKESEFLAPP RLQNAEFIGQ QPKLTVDATV LADGAEVVCA FINDDLSAPV 60
LEKLAANGTK LIALRSAGYN HIDLEAAQRL NLAVVRVPSTY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVNKTGVVV GTGQIGATAFA RIMAGFGCKI LAYDPYPNPQ 180
VITALGASYLE LDTLLAQSHI VSLHCPLSPA THHHLINASSL SRMQQGSMLI NTSRGALVDT 240
PAVIEALKRG HLGYLGLDVY EEEADLFFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALDAIAGT TLANIDAWAA GRPQNVLVRLE T 331

SEQ ID NO: 4      moltype = AA length = 331
FEATURE          Location/Qualifiers
source           1..331
               mol_type = protein
               organism = Methylobacillus methanolivorans

SEQUENCE: 4
MRIVFFSTQV YDRESFLAPP KPQNAEFIGQ QPKLTVDATV LAAGAEVVCA FINDDLSAPV 60
LEKLASGGTK LIALRSAGYN HIDLEAAQKL GLAVARVPSTY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVNKTGVVI GTGQIGATAFA RIMAGFGCKV LAYDPYPNP 180
VIALGACYLD LDTLLAQSHI VSLHCPLSPA THHHLINANSL AHMRRGSMLI NTSRGALVDT 240
PAVIDALKSG HLGYLGLDVY EEEADLFFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALDAIAST TLSNIEQWAT GTPQNVLVRLE A 331

SEQ ID NO: 5      moltype = AA length = 331
FEATURE          Location/Qualifiers
source           1..331
               mol_type = protein
               organism = Methylobacillus arboreus

SEQUENCE: 5
MRIVFFSTQV YDRESFLALP RSPNTEFVFQ QPKLTADTAV LAAGAEVVCA FINDDLSAPV 60
LFTLAAGGKT LIALRSAGYN HIDLEAAHTL KLEVVRVPSTY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVNKTGVVI GTGQIGATAFA RIMAGFGCKV LAYDPYPNP 180
VEALGVRYLE LDELLAQSQI VSLHCPLSPA TYHLINEASL NRMQRGSMLI NTSRGALVDT 240
PTVIDALKSG HLGYLGLDVY EEEADLFFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
HEALDAIAST TLDNIVAWAE GRPQNVLVKLE T 331

SEQ ID NO: 6      moltype = AA length = 331
FEATURE          Location/Qualifiers
source           1..331
               mol_type = protein
               organism = Methylobacillus caricis

SEQUENCE: 6
MRIVFFSTQV YDRESFLEHD KPNDAEFVFQ QPKLTLETAV LAAGAEVVCA FINDDLSAPV 60
LEQLAANGTR LIALRSAGYN HIDLEAAATRL GLAVARVPSTY SPHAIAEHTV ALILALNRHL 120
TRAYNRTREG DFSLRGLTGF DLVNKTGVII GTGQIGATAFA KIMAGFGCKL LAYDPYPNP 180
VVKLGQYLED LDELFQQAQI VSLHCPLSPS TYHLINEASL GRMQSGSMLI NTSRGALVDT 240
PAVIEALKSG HLGYLGLDVY EEEADLFFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALNIAST TLANISAWAE GSPQNVLVIE G 331

SEQ ID NO: 7      moltype = AA length = 331
FEATURE          Location/Qualifiers
source           1..331
               mol_type = protein
               organism = Methylobacillus gramineus

SEQUENCE: 7
MRIVFFSTQT YDRESFLGLD KPTGAEFVFQ QPKLTVDATV LAAGAEVVCA FINDDLSAPV 60
LDQLAATGTR LIALRSAGYN HIDLEAAQRL GLVVTRVPSTY SPNAIAEHTI ALILALNRHL 120
TRAYNRTREG DFNLRGLTGF DLANKTGVII GTGQIGATAFA RIMAGFGCKL LAYDPYPNP 180
VITALGAEYLP LDALFSRAQI VSLHCPLSPN TFHLINAETL NRMQRGSMLI NTSRGALVDT 240
PAVIEALKSG HLGYLGLDVY EEEADLFFFED LSDFPLQDDV LARLLTFPNV IITAHQAFLT 300
REALHAIAST TLANISAWAA GSPCNLVKA E 331

SEQ ID NO: 8      moltype = AA length = 330
FEATURE          Location/Qualifiers
source           1..330
               mol_type = protein
               organism = Methylobacillus glycogenes

SEQUENCE: 8
MQIVFFSTQV YDRESFLSQP ALSDCELVFQ QPKLTVDATV LAAGAEVVCA FINDDLSAAV 60
LEKLAAGGTR LIALRSAGYN HIDLEAAQRL ALAVVHVPSTY SPHAIAEHTI ALILALNRHL 120
NRAYNRTREG DFSLRGLTGF DLVNKTGVII GTGQIGATAFA RIMAGFGCRL LAYDPYPNP 180
VIALGAEYLP LEQLFPQAHI VSLHCPLSPA THHHLINPQTL ALMQKGSMLI NTSRGALVDT 240
PAVIEALKTG HLGHHLGLDVY EEEADLFFFSD LSDYPLQDDV LARLLTFPNV IITAHQAFLT 300
REALAGIAST TLLNVAAWRR GQPQNVLVQPG 330

```

-continued

```

SEQ ID NO: 9          moltype = AA length = 329
FEATURE             Location/Qualifiers
source              1..329
mol_type = protein
organism = Escherichia coli

SEQUENCE: 9
MKLAVYSTKQ YDKKYLQQVN ESGFGELEFF DFLLTEKTAK TANGCEAVCI FVNDDGSRPV 60
LEELKKHGVK YIALRCAGFN NVDLDAAKEL GLKVVRVPAY DPEVAEAEHAI GMMMTLNRR 120
HRAYQRTRDA NFSLEGTLGF TMYGKTAGVI GTGKIGVAML RILKGFMRL LAFDPYPSAA 180
AILELGVEYVD LPTLFSESDV ISLHCPLTPE NYHLLNEAAF EQMKNGVMIV NTSRGALIDS 240
QAIAEALKNQ KIGSLGMDVY ENERDLFED KSNNDVIQDDV FRLSACHNV LFTGHQAFLT 300
AEALTSISQT TLQNLNSNLEK GETCPNELV                                         329

SEQ ID NO: 10         moltype = AA length = 335
FEATURE             Location/Qualifiers
source              1..335
mol_type = protein
organism = Fusobacterium nucleatum

SEQUENCE: 10
MQKTKIIFD IKDYDKEFFK KYGADYNFEM TFLKVRLEE TANLTKGYDV VCGFANDNIN 60
KETIDIMAEN GIKLLAMRCA GFNNVSLKD VNERFKVVRPV AYSPHAIAEY TVGLLILAVNR 120
KINKAYRTR EGPNFSINGLM GIDLVEKTAG IIGTGKIGQI LIKILRGFD M KVIAYDLFPN 180
QKVADELGEF VYSLDELYAN SDIISLNCP TLKDTKYMIRN RSMLKMKDG V ILVNTGRGML 240
IDSADLVEAL KDKKIGAVAL DVYEEENVF FEDKSTQVIE DDILGRLLSF YNVLITSHQA 300
YFTKEAVGAI TVTTLNNIKD FVEGRPLVNE VPQNQ                                         335

SEQ ID NO: 11         moltype = AA length = 329
FEATURE             Location/Qualifiers
source              1..329
mol_type = protein
organism = Pseudomonas aeruginosa

SEQUENCE: 11
MRILFFSSQA YDSESFQASN HRHGFEHQF QAHLQADTA V LAQGFVVCA FVNDDLSRPV 60
LERLAAGGTR LVALRSAGVN HVDLAAAEL GLPVVHPAY SPHAVAEHAV GLIITLNRR 120
HRAYNRTREG DFSLLHGLTGF DLHGKRVGVY GTGQIGETFA RIMAGFGC EL LAYDPYPNPR 180
IQALGGRYLA LDALLAESDI VSLHCPLTAD TRHLIDAQRL ATMKGAMLI NTGRGALVNA 240
AALIEALKSG QLGYLGLDVY EEEADIFFED RSDQPLQDDV LARLLSF PN VVTAHQAFLT 300
REALAAIADT TLDNIAAWQD GTPRNVRV                                         329

SEQ ID NO: 12         moltype = AA length = 334
FEATURE             Location/Qualifiers
source              1..334
mol_type = protein
organism = Luteitalea pratensis

SEQUENCE: 12
MDIAVFDTRA YDRHAPDAAN TRYGHRELFF EPRLSSRTAT LASGVPVVCP FVNDRADAET 60
IGALHRCGTR LLALRSAGFN HVDLDAAAAI GMRVVRVPEY SPYAVAEELAM CLVLTNRKV 120
HRAYNVRREA NFSLEGVLGF DLHGKTFGII GTGRIGRVLIA RIAHGFCCR LATDASPDEA 180
LQRELGVVEY EASTLYREAD VISLHVPLTT ETHHMIDKDA FAQMKGAVML INTSRGALVD 240
ARALIQALKS RHVGAAGLDVY YEEEGPLFFR DLSAQVLQDD VLARLLTFPN VLITSHQGFL 300
TREALANIAD TTLASVSAFA RGEALVHEIG QRVP                                         334

SEQ ID NO: 13         moltype = AA length = 331
FEATURE             Location/Qualifiers
source              1..331
mol_type = protein
organism = Streptomyces rubrolavendulae

SEQUENCE: 13
MDVLAFCVGQT DEKPILLERAF AGHHGLHCLD VFLTEDTAPL AAGHEAVSTS VNADLGARVL 60
RTLAAGGTRL VAQRSTGFNN IDLDVAREL MGTVARVSYY PYSVAEFWLT LAMAVNRRV 120
RAAHTRTRDFD FRLDGLLGRD MRGRTVGVLG TGKIGEAPTR IAHGFGMRLL GWDLAENPAC 180
RELGMayVGK ELLAASDLV SLHVPLTEKT WHLDDADALA AMKDDAILVN SSRGGLVDTD 240
ALVAALRAGR FAVGQLDVYE AEAGLFHVDR SLTGVDLDTL ARLMTFPNVL VTSHQAFYTQ 300
EAVGEIVAAT LRNVTDYAEQ RRSENVLVPA G                                         331

SEQ ID NO: 14         moltype = AA length = 334
FEATURE             Location/Qualifiers
source              1..334
mol_type = protein
organism = Aquifex aeolicus

SEQUENCE: 14
MNVLFTSVPO EDVPFYQEAL KDLSLKIYTT DVSKVPENEL KKAELISVVF YDKLTEELLS 60
KMPRLKLIHT RSVGFDHIDL DYCKKKGILV THIPAYSPE VAEHTFAMIL TLVKRLKRIE 120
DRVKKLNFSQ DSEILARELN RLTGIVGTG RIGSRVAMYG LAFGMKVLCY DVVKREDLKE 180
KGCVYTSLDE LLKESDVISL HVPYTKETHH MINEERISLM KGKVYLINTA RGKVVTDAL 240
YRAYQRGKFS GLGLDVFED EILILKKYTC GKATDKNLKI LELACKDNVI ITPHIAYYTD 300

```

-continued

KSLERIREET VKVVKAFVKG DLEQIKGNFV VGPS	334
SEQ ID NO: 15	moltype = AA length = 310
FEATURE	Location/Qualifiers
source	1..310
	mol_type = protein
	organism = Armatimonas rosea
SEQUENCE: 15	
MKIIIVTAAYS DLEEEYDHMV RLLPEHDVAP TLAPDAPVLC VTSRDLVTLE LLDSLPGVRL	60
IATRSTGFEH IDLAALARARG ITVCNVPHYG DHTVAEFAFG LLLTTLKRHLH TVLKDADAGR	120
FPPGGLRGLD LFGLTLGVVG AGRIGRCAA K IGVGFGMTVL AHDIAPLPDE AERIGFTYVG	180
LDELLTRADV LTLHTPLTPQ TYHILGAAEL ARLPRAVAVLI NTARGSVVAL AAVLQALDSG	240
QLAGAALDVL EDEERIAELA PPLLERYPNL LLTPHVAYHT VGSVGRIRQT TADNIRAFLN	300
GAPQNTVPLP	310
SEQ ID NO: 16	moltype = AA length = 347
FEATURE	Location/Qualifiers
source	1..347
	mol_type = protein
	organism = Neosartorya fumigata
SEQUENCE: 16	
MKLAVFSAKS YDKLFLDSSL EKDYASFCEI VYHSFALSSE TVSLAKGSDA VCAFVNLDL	60
APVLKSLRCA GIRAILLRC GFNNVNDLQVA EELGFFVANV PSYSPEAVAE FAVALIQTLN	120
RKTHRAYNRV REGNFNLEGF LGHTLHGKTV GIVGVGRIGL ALAKIFHGFG CRLLASDPFG	180
GEEFRKYGEF VELSELLAQD DVVSLHCPLT ESTRHIINDE TLAQMRKGAL LVNTSRGGGL	240
NTKAAIRALK EGHLGGALD VYEGERGSLFY NDHSTEIIHD DTLMRLMTFP NVLVCGHQAF	300
FTQEALCEIA GTVTLGNLQDF VLKRTCKNSL VREGHLLVRR DTEPVRL	347
SEQ ID NO: 17	moltype = AA length = 340
FEATURE	Location/Qualifiers
source	1..340
	mol_type = protein
	organism = Flavobacterium sp.
SEQUENCE: 17	
MTKINLLSHK ISFFSTQPYD QTFFEKEYNAD FGFELEEFFDI QLNEQTVTLV DNKAIVCVFV	60
NDVVNAAVIK KLAKEEVKII ALRCAGFNNV DLEAAKKHG I QVCRVPAYSP EAVABHAMTM	120
ILTLNRKTHK AYNRVREQNFI SLHGLLGFDL HGKTIGIIG GNIGKAFSKI AKGFGSTVVA	180
YDVVKDSGME ADGVQYVALD TLIKKSDII S LHCPLNDETR HLINKKTLAQ MKTNVMIINT	240
SRGALINTSD VIKGLKSGKV GYLGIDVYEQ E EKLFIFRDL DDIIIQDDKIQ RLMSFPNVLV	300
TAHQAFFTNE ALTQIALITL SNVKELALNG QTKDQLTQLV	340
SEQ ID NO: 18	moltype = AA length = 260
FEATURE	Location/Qualifiers
source	1..260
	mol_type = protein
	organism = Thanatephorus cucumeris
SEQUENCE: 18	
MRCAGYNVND LKVAKELGIL VVNVPAYSPE AVAEFAVGML LTIVRKYHKA YNRVREGNFL	60
LDGLTGPNLE GKTIGILGTG KIGLCTGRIL SHGFRAKVG YDPYPNPTAA EENGIKYVSL	120
DELFKTSVDI SLHCPLTPET KYIVNEEALK TTKPGLIIVN TSRGALINTS DLIHGLKSGH	180
VGAVGMDVYE RESKYFGRDS SNKIIQDDQL SRLVSFHNVF ISGHQAFLTQ EALSAIARVT	240
VENLRLLEEG SPCSNIVTEG	260
SEQ ID NO: 19	moltype = AA length = 337
FEATURE	Location/Qualifiers
source	1..337
	mol_type = protein
	organism = Candidatus Heimdallarchaeota archaeon
SEQUENCE: 19	
MKIAFFEVKS WEIDYLTKLFR TEEDFDLTF TDQPLNQKNV ENFTDINIAS VFIYSTINKS	60
VIKSLNLNLKF IATRSTGFDH IDLTECDKNK ILISNVYPYGG ENTIAEHTFG LILALSRIH	120
KSYVRTQRND YSIEGLKGFD LQGRTLGIIG LGRIGMHVAR IARGFGMKVL AYDIQDSFF	180
SDLINFTYTS FEEVLLKLSDI VTLHVPYNNK THHLINKNTI KLFKKGAVLI NTARGGVET	240
ESLLEALENK TLSGIGLDVI EGEETYIKEEK QLLYDSEKID IWKKIIQDHI LLKKDNVVFT	300
PHIAFYSQEA LERILQTTKD NIKGFTSNRP QFIVTNK	337
SEQ ID NO: 20	moltype = AA length = 337
FEATURE	Location/Qualifiers
source	1..337
	mol_type = protein
	organism = Candidatus Hydrogenedentes bacterium
SEQUENCE: 20	
MSNIAIYDSK PYTRHSEAR NTGLGPITFF EHKLTDGTAG SAAGFDAVCI FVNNTVDAEV	60
AMRLKELGVG LVALRCGYN NVDLNACRH GISVVRVPGY SPPSVAEHSI ALMLALNRHI	120
MRRAHARVREG NFSLNGLTGF EMRGKTAGIV GTGRIGTCAA EILSGFRCSI LAYDPHPAEA	180
LSKNPRLRYT DMDTLLHQSD IISLYVPLLP TTRHLINDEA IRKMKRGVML INTSRGALVD	240

-continued

TAALIRGLKS GKGGSAGLDV YEEESDYFFE DRSEYIIADD LLARLLTFNN VLITSHQGFL	300
TNEALDAIAL TTFQNIQEYL EGKREKELSN GICPKCG	337
SEQ ID NO: 21	moltype = AA length = 274
FEATURE	Location/Qualifiers
source	1..274
	mol_type = protein
	organism = Microgenomates bacterium
SEQUENCE: 21	
MDIGFFEVND WEVPIIQLKH PFAFISSEKL VAENAKKYKN LQVNSTFIYS QLNSSVLVQM	60
PKLRCIATRS TGFDHVDAKY CARNNIVVCN VPRYGDRTVA EFTFALILSL TRRIYESVAL	120
TRLGSFDNHG LTGMDELEHKT IGIIGLKGIG QEVAKIAARAF NMKILAFNRS HDDTLAAELG	180
LEYVELNELL KRSDDVVTLHL PLLPTTHFI NKNILTMKKG SYLINTARGG LVETQALIIG	240
LHKGILAGVG DVLEEEESVL NEEAELISSE FRKK	274
SEQ ID NO: 22	moltype = AA length = 328
FEATURE	Location/Qualifiers
source	1..328
	mol_type = protein
	organism = Candidatus Nanohalobium constans
SEQUENCE: 22	
MTEVKVAAFWD TDENWEKEYLE NKDFPFEIKP FEESLTKDIA AKTEGFDAVS VFVSSQVSSE	60
VIEELDVII ACRSTGFDHV DIETASEQGI KICNVPRYGA NTVAEHAFLG ILTLSRRRIHE	120
AISKEVNGEF DHQGLKGFDL EGKTLGVVGT GSIGQKVIEA AHGFNNMDVLA SDPYPKEGLE	180
EELGFMVYVM EDLIKKSDII TIHCPLTDEN HHLFSEEQFK LMDETVLINT ARGELVDTTA	240
LIRALENGNV SRAGLDVLEE ECVIEEDIEI LGDISDECOP QKIVEDHILM ERDDVVVTPTH	300
NAFNSREALQ RIEDKTLENI EKGRNTVN	328
SEQ ID NO: 23	moltype = AA length = 332
FEATURE	Location/Qualifiers
source	1..332
	mol_type = protein
	organism = Candidatus Peregrinibacteria bacterium
SEQUENCE: 23	
MKIAFLEVSK SDKIFLRRRL PDDELIFINE NLNEENIKKI KDCEVLSVFI YSQVNKKNLD	60
KLLKLKLITT RSTGFDHIDA QTCKNKGISI CNVPFYGENT VAEHTFALIL ALSRNWKSY	120
LRTLRRDYST TGLMGFDLKG KTLGVIGAGR IGLHVIRIAK GFGMVNLAFD LNRDKFISEI	180
LDSYGTIEE VLKASDIITL HVYPNKYTHH LINKDNLRV KKGAILINTA RGIIETQAI	240
IEALDKGLLS GVGIDVLEGE EMVLEEKKLN FDEKNLESMS HLIQDHILLS RDNVVFTPPI	300
GFYSKEALER ILETTVESIE GFKKGNLVNS IS	332
SEQ ID NO: 24	moltype = AA length = 347
FEATURE	Location/Qualifiers
source	1..347
	mol_type = protein
	organism = Candidate division BRC1 bacterium
SEQUENCE: 24	
MTRKIAFPDA KPYDIKSFT MNKDPGFADIA YFETHLNQAT ASLSGDFDAV CAFVNDTIDK	60
AVIDEIIKHG VKLIALRCA G YNNVDFKAAY KKIHIARVPA YSPYAVAEEHA AALMPSLNRK	120
THRAFFRTRD GNFSLNGLLG FDMHGKTAGV IGTGKIGKCL IAILKGVGMK LLAYDPFPDE	180
AFAAADAGVQY VSLDEIYQQA DIIISLHCPLN KDTTEYMINAQ AISRMKNGVM IIINTGRGKLI	240
HTQUALIDGLK SGKIGAAGLD VYEEESEYFF EDKSAQMITD DILARLLMFP NVLVTSHQAF	300
FTEKEALRNIA QTTLQNCQDF FEHEKLVNEI CYKCGGPCTR KEKKLCF	347
SEQ ID NO: 25	moltype = AA length = 327
FEATURE	Location/Qualifiers
source	1..327
	mol_type = protein
	organism = Methanomethylophilus alvus
SEQUENCE: 25	
MVKFVNYR PDEEKYFQQY AEEFGFELGY TEESPTIDNC GLADGCFD FIS IITTPITPEM	60
MDRFKEGGVR MISTRIGYD HIDIAYAKRI GMTVTHITYD PEGVAEYTVM MMLMAIRKAK	120
NILERGKDND PTLDGLIGGE LGDMSVGIIG TGRIGRVSRLR DLSGFGCR LY YCNRSPEEA	180
DRYAERVDMD ELLGMCDVVS LHLELNKDTF HIMGPEAISK MKKGALIVNT ARGPLIDTDA	240
LAAGLDSGQV GGAALDVIED EFGLYYYDCS GRELDNKALK VLRLGRPNVLI TQHMAFYEN	300
AIRDMVYNSL YGMKMLADGK EIPYRLA	327
SEQ ID NO: 26	moltype = AA length = 339
FEATURE	Location/Qualifiers
source	1..339
	mol_type = protein
	organism = Candidatus Wolfebacteria bacterium
SEQUENCE: 26	
MKIAFFQIED WEIEHIIKEQL AGHELFFSK E KLSAEALPEQ RDFDIVSVFV GSKIDQAVLA	60
ALPNLKLVTT RSTGFDHVDL PMAQSMNIAT GYVPGYGDNT VAEFAFGLIL ALSRKIYDSV	120
DRLRETGVYS YVGLRGFDLQ GKTIGVMGTG RIGQHVRIRIA KGFGMQVIAF DAFTPKAELAT	180

-continued

ELGFVEYVGLD DLLGRSDVIT IHVPYLPSTH HLINMDSLTK IKKGALLINT ARGAVVETDA	240
LVKGLADGIL CGAGGLDVLEE EGVLADEKTL VLGHPEEHN LKTVLENHVL IDMPNVLVTP	300
HNAFNTQEAL QRILDTDICN IQEFLGKGIA KCPIPAPKA	339
SEQ ID NO: 27	moltype = AA length = 202
FEATURE	Location/Qualifiers
source	1..202
	mol_type = protein
	organism = <i>Candidatus Anoxychlamydiales bacterium</i>
SEQUENCE: 27	
MKIAIFGSQE YDIESLNEAN KKYSNDLTFF KAYLNEKTAS LAKNFDAVCL FVQDTANKAV	60
IDILAKQGVK TIALRCAGFN NVLDLSAKTN NISVNVNPY SPHVAEAEHAV GLILSLNRKI	120
HHAYTRIRDG NFSLGKFLGF DLYGKTVGVI GTGKIGQETA KIFKLGFGCK VIASDPYPNK	180
TLEKECQSPL PKGRGLNRED QG	202
SEQ ID NO: 28	moltype = AA length = 453
FEATURE	Location/Qualifiers
source	1..453
	mol_type = protein
	organism = <i>Monoraphidium neglectum</i>
SEQUENCE: 28	
MRAPAAQQPRA AGAARGPATA PRAPALRGAT WRGACRVTQ VDAVEAPRPA DPVEAPKPS	60
DPELAAKRSA PKIMTFSAQP YVNKYFLEPL QAAGFDNIKG TEASAGTGAP RACGRACTQR	120
QROPPPPAEPG HASARLGI DT VKIASGYDAI CLEFVNDLCDA EVINALAEGG VKFIAMRCAG	180
YDRVDVAAAN AAGIKVVVRP TYSPTSVAEH AVALLFAIDR HIVGAINRTV QGNYSLSGLV	240
GRQLGGKTIG ILGTGAIGAE AARIKGIGM KVLAYDIRPN PAVEALGIPY LPWEELPQA	300
DVISLHMPLL PSTYHPIDEA KVAMMKPGVV LLNVSRGGLI DTPALIYGVR SGHIGGAGLD	360
VYEKEGERPC PRAARGVALR DLTDLQPRDR MVPWDRLRIE LIALPQVLVT PHTAPLTAEA	420
LVNIADTTVQ NIDQFMLGKP LGNEVKPPPPP AKS	453
SEQ ID NO: 29	moltype = AA length = 482
FEATURE	Location/Qualifiers
source	1..482
	mol_type = protein
	organism = <i>Rattus norvegicus</i>
SEQUENCE: 29	
MLLRVATQRL SPWRGYCSRG SQGGLSQDFV EALKAVVGSP HVSTASAVRQ HHGHDESMHR	60
CRPPDAVWWP QNVDQVSRLA SLCYNQGVPI IPPGTGTGVE GGVCAVQGGV CISLTHMDQI	120
MELNTEFDLSEV VVEPGVTRKA LNTHRLRNSGL WFPVDPGDA SLCGMAATGA SGTNAVRYGT	180
MRDNVINLSEV VLPGDRLLHT AGRGRHYRK5 AAGYNLTGGLF VGSEGTLGII TSATLRLHPA	240
PEATVAATCA FPPSVOAAVDS TVQILQOAAVP VARIEFLDEV MMDACNRHSK LNCPVAPTLF	300
LEPHGSQCAL AEQLQRTEAI TQDNGSHFS WAKEAEKRNE LWAARHNawy AALALRPGSK	360
AYSTDVCPVI SRLPEILVET KEELKASKLT GVIVGHVGDG NFHCILLVNP DDVEEQRRVK	420
AFAENLGRRA LALHGTCTGE HGIGLGKRLQ LQEEVGPVGV ETMRQLKDTL DPRGLMNPKG	480
VL	482
SEQ ID NO: 30	moltype = AA length = 323
FEATURE	Location/Qualifiers
source	1..323
	mol_type = protein
	organism = <i>Desulfurispira natronophila</i>
SEQUENCE: 30	
MTDVFFYEAF AEEAEQLQRF LPNSVSAGFT DRTIQESGD HPPARVISVR TQSAIPPSWS	60
PHIEAILTRS TGFDHIQRYF WETGQKIAAG YLPLYCNRSV AEQAMLLWMA LLRQLPRQME	120
QFGSFHARDGI TGGEAQGRTL MVAGVGNIGH EICRIGSGLG MTVLGVIDE RHDDVLYVSP	180
HEAIQQAHHV VCSMCLTNEN RHYFHYDMLR QARPGAVFVN IARGELSPA DLMRALDESI	240
LAGVALDVFD EEKELATALR QGSPLVQPTV CQTLERLARRS NVIMTPHNAF NTVESVERKS	300
RQSLEQLQHF FATGEFLWPI GRN	323
SEQ ID NO: 31	moltype = AA length = 473
FEATURE	Location/Qualifiers
source	1..473
	mol_type = protein
	organism = <i>Aeropyrum pernix</i>
SEQUENCE: 31	
MARIAEELEK IFGPEKVVSD PHIVRLYSRE PSGLEGRAEVA VVFPESAQDV SRLVRYAYS	60
EVYIYPQGSS TDLAGGAFPE RPGVVVSMER MRRVREHSV DSVAVVEPGV RLWDLNVELS	120
KYRYMFPIDP GSVKVATVGG AINTGAGGMR GARYGTMRDW VLGLEIVLPD EEGTILRVGC	180
RTLKCRQYD DARLIVGSEG TLAIVTEAIL KITPMPENVV VVLAFFPTLQ QLVDAVIEVK	240
SRAIDTLLME FMDVDSARLA AETLGAAIRP DGHMILLVGVVN REASTRVL EEMVSIAKAA	300
GAASVYTAKS MEEAEKKLL EIRRSLSFATQ ALLTQKQPKG RKVMMMLMEDI AVPPSKLLDA	360
VERLKELEAK YGFKTVLGGH IGDGNLHPTI SYPVDDKEAK EAALKWYYDV MRMAIELGGT	420
VSAEHGIGVL KKEALRLELE RMGSVKALEI MAGIKRVPDP KGILNPGKVV AAE	473
SEQ ID NO: 32	moltype = AA length = 333
FEATURE	Location/Qualifiers

-continued

```

source          1..333
               mol_type = protein
               organism = Synechocystis sp.

SEQUENCE: 32
MKIAFFSSKA YDRQFFERAN HHHGREMLFF DAQLNLDTAI LAEDCPVCL FVNDQAPAPV 60
LEKLAAOGTT LIALRSAGYN NVDLKTTAAAL GLKVVHVPSE SPHAVAEHTV GLILALNRKL 120
YRAYNVRVDD NFSLEGLLGF DLHGSTVGI GTGKIGLAFA KIMNGFGCHL LGYDAFPNEN 180
FTAIGQAIYV PLNELLAQSD IISLHCPLLP ETHYLINSST IARMKPGVML INTSRGHЛИD 240
TQAVIQGIKS HQIGFLGIDV YEEEEALFFT DHSDTIIQDD TFQQLQSFPN VMITAHQGFF 300
THNALETIAE TTLANIAEFE QNQPLTYEV CPH                                333

SEQ ID NO: 33      moltype = AA  length = 349
FEATURE           Location/Qualifiers
source            1..349
               mol_type = protein
               organism = Elusimicrobia bacterium

SEQUENCE: 33
MERIKLAFFD AKPYDIEFFN IENAKYGYHI KYFLNRLNSD TVAMSQGFNA VCIFVNDNVT 60
KDVVKILLKN KIIKIALRCA GYNVNVDLSSV YNKIHVVVRVP DYSPYVAEAEH TVGIMLSSLNR 120
KIHKAYYRTR DNNFNINGLL GFDMNGKTAG IIGTGKIGKV LIKILKGFnV KILAYDPYPD 180
KDFAEKMEIS YVPLDKLYRS SDIISLHCPL TKDTFHLIDH SAISKMKTGV MVINTGRGRL 240
IDTKALTEAL KHGKIGSAGL DVYEESEYF FEDFSDKILP DDALARLLSF NNVLITSHQA 300
FFTKEALTN ARTTLGNIFA YSKGEKLANE VCYKCDKRSC CSGKSGKCF                349

SEQ ID NO: 34      moltype = AA  length = 310
FEATURE           Location/Qualifiers
source            1..310
               mol_type = protein
               organism = Methanoculleus bourgensis

SEQUENCE: 34
MVVKLVFFDL DEHEQDVVRD AFASESGYEV DLHSDALTRO NVSLARDADG IAVFVTSHVT 60
REVLDRLPNL KVIAAMSTGF DHIDVDTCRE RKTAVCNVPL YGDTTVAEFA FGLILALARK 120
FRPTFARTAR GDFSRGTGLQ MDLAGKTGL IGTGRIGSHL ARLAHAAGMN VVAYDLHPNL 180
ALTKDYGVRVY LDLLDDLREA DVISLHVPT KATHHLINAD RLRLVKDTAL LVNTSRAGVV 240
DTQAVAAALR EGRLGGGRTR YLRRGGADLDR GGVFWGRRAL GGRTQGCARE LRHPAIGPGN 300
TLAAQRQLQYP                                         310

SEQ ID NO: 35      moltype = AA  length = 340
FEATURE           Location/Qualifiers
source            1..340
               mol_type = protein
               organism = Dictyostelium discoideum

SEQUENCE: 35
MKITLFSSKP YWVKWFNELN KFSYEINYVT SACDIKSVNE AKGSEAVCCF VNDDLSKEVI 60
ETLHSNGTKV ILMRCAGFNK VDLDTANKLG IPVLRVPAYS PNAVSEYALS LIMALNRKTH 120
KAHDHRVRDAN FEINGMEGFN MVSKVYGYIVG TGNIGEQLCR VLKLGFGAKV IAYDIIENKA 180
VTDIGIEYVK TLDEIWKQCD VISLHTPLNS QTKYMVNSES IEKMRDGVMN INVSRGALVN 240
ASDAIVGLKS GKISSLGMVD YENETDYFYQ DHNGSIKKD NLSSLISYPN VMITSHQAWY 300
TKEAISCICG TSLQNFVDFR SNQIKKSNLV NNPPISSQPTQ                340

SEQ ID NO: 36      moltype = AA  length = 333
FEATURE           Location/Qualifiers
source            1..333
               mol_type = protein
               organism = Lactobacillus delbrueckii subsp. bulgaricus

SEQUENCE: 36
MTKIFAYAIR EDEKPFLKEW EDAHKDVEVE YTDKLLTPET VALAKGADGV VVYQQOLDYTA 60
ETLQALADNG ITKMSLRNVG VDNIDMAKAK ELGFQITNVP VYSPNAIAEH AAIQAAIRL 120
QDKAMDEKVA RHDLRWAPTI GREVDRQVVG VIGTGHIGQV FMQIMEGFGA KVIAYDIFRN 180
PELEKKGYVV DSLDDLYKQA DVISLHVPDV PANVHMINDE SIAKMKQDVF IVNVSRGPLV 240
DTDAVIRGLD SGKIFGYAMD VYEDEVGIFN EDWEGKEFPPD ARLADLIARP NVLVTPTAFA 300
YTTHAVRNMV VKAFDNNLEL VEGKEAETPV KVG                                333

SEQ ID NO: 37      moltype = AA  length = 332
FEATURE           Location/Qualifiers
source            1..332
               mol_type = protein
               organism = Fusobacterium gonidiaformans

SEQUENCE: 37
MIKMRVLFFD AKSYDKENFD AYKEKYGF DI KYLKVKLNNE TVDFVKGYEI ISIFVNDTVN 60
PPVIDKLIY GVKLIVLRC GYNVNVDVNYI NGRIKLVRPV AYSPYSVAEY TASMVMTLNR 120
KIHKAYYRTR EGNFSINGLM GFDLHKKTVG VIGAGRIARI FIKIMRGFDA RVIAYDPYPN 180
ESPARDLGYE YVDLDTLYRE SDIISLHCPL TRENTYLINR ESMKKMKDGV MIVNTGRGRL 240
IDTIDLIEAL KDKKVGAAL DVYEEEAGYF FEDMSSSIIE DDILGRLLSF NNVLITSHQA 300
YFTKEAFRDI TLTTLHENIQS FLQGKELENE IK                                332

```

-continued

SEQ ID NO: 38	moltype = AA length = 360
FEATURE	Location/Qualifiers
source	1..360
	mol_type = protein
	organism = Kiritimatiella glycovorans
SEQUENCE: 38	
MSAPSILFFD AKSYDREFFQ	AANGNFGYEL TFLGHRLTGE SARLAEGHDA VCAFVNDRLS 60
PEVADTLADS GIGLVALRCA	GYNVNVDLRSF FGRIHVVRVP AYSPhVAEh TLALILSLNR 120
KTHRAYYRTR DGNFAIEGLL	GFDMAGKTAG VVGTGSIGRC AAQLLRGFGM EVLAFDVNP 180
LEWAERSGVT YVDMGELYR	SHLTLHCP L TSDNHHMINA DTIGQMRDGv MIINTGRGGL 240
IDAPALTEAL KDRKVGAAGL	DVYEEEDQYF FEDFSVRGLD DDVLARLLTF PNVLITSHQA 300
FFTREALTNI ADTTLGNIRD	YFEGRPLENE ICYRCCDEGC RKEEKGRCFE SGPEsAPPSE 360
SEQ ID NO: 39	moltype = AA length = 363
FEATURE	Location/Qualifiers
source	1..363
	mol_type = protein
	organism = Lentisphaerae bacterium
SEQUENCE: 39	
MQDKIRIAFF DTKPYDRLSF	DAIKERYHVE IVYFETRLLTP ASAHLAEGFD AVCAFVNDDL 60
CAETIHTLEA LNRVLRIAMRC	AGFNNVNDLAA ACHRMKVVR PKYSPYAVAE YTIGLLLCLN 120
RHLHRAPCRI RENNFSINGF	MGFDLHGKTF GIIGTGKIGL TFAQILGGFG VRILASDPFP 180
NEKAAAELKM EYVPFETLYR	ESDVISLHCP LTPENHHMIC SDTIGQMKTG VFILNTSRGK 240
LIDTSDLIEA LKSGKVGGAG	LDVYEEEADY FFEDRSDRWI NDDVLARLMT FPNVLITSHQ 300
AFFTREAVNN IAEITMDNIR	DFQEGRRELVN DVCFHCDGTR PCCHAKENTA TSNPETSGSS 360
ASR	363
SEQ ID NO: 40	moltype = AA length = 332
FEATURE	Location/Qualifiers
source	1..332
	mol_type = protein
	organism = Stieleria varia
SEQUENCE: 40	
MKIAVFSTKP YDESFLRAAS	ENHSHTLTFL EPRLTAKTAA LANGFDAVCA FVNDQLDSDV 60
IHTLSDHGVRI AIALRCAGFN	NVDLAAASEC GVTVVRVPAY SPHAVAEHTA ALVLALNRKI 120
HRAYTRVRDG NFALGGLLG	DLNLRVGVV GTGKIGEIFA KIMSGFCCKL LGFDVQQNPE 180
CERLGGMQYVS LDELFAESDV	ISLHCPLTPQ THHLIDAAAI QKMKPGVMIV NTSRGAVIDT 240
GAVINGLKSG KIGHLGIDVY	EEEADLFFFED LSSQVIPDDM LSRLLLTFPNV IVTGHQGFFT 300
EEALSCIAET TLQNLTDLES	TGNCNPNAVTT TG 332
SEQ ID NO: 41	moltype = AA length = 329
FEATURE	Location/Qualifiers
source	1..329
	mol_type = protein
	organism = Escherichia coli
SEQUENCE: 41	
MKLAVYSTTKQ YDKKYLQQVN	ESFGFELEFF DFLLTEKTAk TANGCEAVCI FVNDGSRPV 60
LEELKKHGVK YIALRCAGFN	NVDLAAKEL GLKVVVRVPAY DPEAVAEHAI GMMMTLNRRI 120
HRAYQRTRDA NFSLEGLTGF	TMYGKTAGVI GTGKIGVAML RILKGFGMRL LAFDPYPSAA 180
ALELGVEYVD LPTLFSESDV	ISLHCPLTPe NYHLLNEAAF EQMKNGVMIV NTSRGALIDS 240
QAIAIEALKNQ KIGSLGMVDV	ENERDLFFFED KSNDVIQDDV FRRLSACHNV LFTGHQAFLT 300
AEALTSISQT TLQNLNSLEK	GETCPNELV 329
SEQ ID NO: 42	moltype = AA length = 340
FEATURE	Location/Qualifiers
source	1..340
	mol_type = protein
	organism = Galdieria sulphuraria
SEQUENCE: 42	
MSKSLFRVNV FSTKNYDQIG	LTEALGRSSD AAPISFRFLS SRLSASTAEL AKDAEAVCVF 60
VNDTVNSEVL KILHDKGQL	I ALRCAGFN VDLKAASELG ISVVRVPAYS PNAVGFEVAA 120
QLLTLARKIH KAYNRVRENN	FDLQGLVGIE IRGKKAGIVG TGKIGVATAk VLKGLGLELF 180
GFDFVFNQEF IDLGGYQLSL	DELLSQSDIV SLHCPLNEST KYLIRSETIA KMKSGAYLIN 240
TSRGALLDTo AVLDALYSGH	IGALAIADVYE GEGDLFFFEDR SGAIVHDPIl AKLQALPNVL 300
VTGHQAFLTD VALRNICNIT	IDNLSNFYYQ RPLKNKVDPK 340
SEQ ID NO: 43	moltype = AA length = 331
FEATURE	Location/Qualifiers
source	1..331
	mol_type = protein
	organism = Treponema pallidum
SEQUENCE: 43	
MRCVVFNLRE EAPYVEKWK	QSHPGVVVDT YEEPLTAKNk ELLKGYEGLV VMQFLAMEDe 60
VYDYMGAACK KVLSRTAGF	DMYNATLLKK HGIRLTVNP YSPNAIGEYA LAAALQLTRH 120
AREIETFVRK RDFRWQPKIL	SKELRCRSRVG ILGTGRIGQA AARLFKGVGA QVVGFDPPN 180
DAAKEWLTYV SMDELLSTSD	VISLHMPATk DSHHLINAKT IAQMKGDVYL VNTARGAVID 240

-continued

SQALLDSDLK GKIAGAALDA YEFEGPYIPK DNGNNPITDT VYARLVAHER IIYTPHIAFY	300
TECTAIENMVF NSLDACTTVL RGEPCAAEIK L	331
SEQ ID NO: 44	moltype = AA length = 313
FEATURE	Location/Qualifiers
source	1..313
	mol_type = protein
	organism = Vitis vinifera
SEQUENCE: 44	
MESIGVLLTC PMNPYLEQEL DKRFLKLFRLW DFPSANDLFR EHSNNSIRAVV GNSFIGADAQ	60
MEALPKMEI VSSFSVGLDK IDLVRCKEKG IRVINTPDVL TEDVADLALA LILATLRRIC	120
ESDRYVRSGS WKKGDFKLTT KFTGKSVGII GLGRIGSAIA KRAEGFSCPI SYHSRTEKPG	180
TNYKYYPSV ELASNCQILV VACALTPETR HIINREVINA LGPKGVVINI GRGLHVDEPE	240
LVSALVEGRL GGAGLDFVFN EPNVPEELLA MDNVVLLPHV GSGTVEVTRKD MADLVLGNLE	300
AHFLNKPLLT PVV	313
SEQ ID NO: 45	moltype = AA length = 345
FEATURE	Location/Qualifiers
source	1..345
	mol_type = protein
	organism = Mycoplasmopsis maculosa
SEQUENCE: 45	
MKIAFFDAKS YDKKYFDLVN KGEHEIVYFE ENLNINNVNL AEGFDAVCCF VNTFGDKYIL	60
ELLSQLGVKV WLQRSMGYNK VDLFAAKELG INVFRVPNYS AESVSEFAVT LLMTLNRVN	120
KAIERVKQYN FNLDGLDGKS IIGSTVGIG AGKIGQGFIR ALKGGMARVL VYDAYSEQNL	180
PNLAIELGFE YAPFTKILRE SDFISHAPS LPSTKYIIDK DAVNLMKKDV VIVNTARGDL	240
VDIKAILYGL ENNIIRGFGA DVLDREEGRF YEDISKNAFE YKQODPEWQK LIEDNRVIIT	300
SHQAFLTDVA LGQIAKILS NASNAENGNF EGALTLLDDG RVLNG	345
SEQ ID NO: 46	moltype = AA length = 349
FEATURE	Location/Qualifiers
source	1..349
	mol_type = protein
	organism = Thermodesulfobacterium geofontis
SEQUENCE: 46	
MIWKIGFFEI EKDWEKNLYL KKFKETLKEK IEDTEIYFYE ETLNENNVR FRDLNIIIIH	60
AESKINSTVV ENLVDTKLLI TRTTGTDHID VLSCEKKGIL VANCPVYAST TVAEHTVALM	120
FALARKLKIA IDKNKKLDFS RDELMGIDLW GKTGVIIGTG RIGSEIARIV YGIGMKILAT	180
DVSPNNEVELV KYVKVYVDL TLLKESDII VMVPYYSQTH HLINRENICK VKEDAIFINT	240
ARGPIVDIEA LIWAKNNKL QGGIAMDVF GERVLMEFQN NLINGAFTAE EYERALKTLS	300
LLNYSNIIFT PHIAYYTKEA MERVIDWWVE NISRFLVCQV LPFQYKFYF	349
SEQ ID NO: 47	moltype = AA length = 343
FEATURE	Location/Qualifiers
source	1..343
	mol_type = protein
	organism = Defluviiotoga tunisiensis
SEQUENCE: 47	
MSKNYKIAIV NSSTFGVYFP DLIQLRKKIG DIERIEVDPN INGKNLAKKL KGFTFVISSV	60
PTPFSEEFFR YNKDVKLIAR HGIGYNNVDI KAATENEVMV TRVYGVHERD AVAEIAVSLM	120
LICLRGIIPA YQAVLQNWKH ERKNDVFGKEL SKITVGIIGY GNIGSRVGEI VKEGPKSEVI	180
AYDPYIADKV IEKTVGQPVF FDELFKKSDV ISLNASLNEG NYHFINKKAF QKMKDGVVIV	240
NTARGETLINL PDFVEALETG KVFAAGLDL ETEPIEPDNP LLKFPNVYIV PHIGGYGTYS	300
LRKMDEKMVE DIEKLVNGEI PDQIVNPEVI SKILETFQGK NNR	343
SEQ ID NO: 48	moltype = AA length = 329
FEATURE	Location/Qualifiers
source	1..329
	mol_type = protein
	organism = Lacunisphaera limnophila
SEQUENCE: 48	
MRVVVFSTKP HDRQFLDAAN AGRHDLVYLE ARLLPETATL AAGAQAACLF VHDHADAABL	60
ATPAGLGVKH LALRCAGFNN VDLAAAARLG ITVARVPPAYS PHGVAEHAVA LFLTLNRRIH	120
RAYNRVRDGN FSLDGLLGFD VHKGTVGVIG TGKIGGLCFAQ IMRGFGCRVL AFDVTRQPAA	180
EALGVEYETL ERLYAESDLI SLHCPPLQT QHLIGAGALA QMRDGVYIVN TSRGPLIDTG	240
AVIDALKSGR LGALALDVYE EEEGVFYEDL SGDILADDQL ARLLSFPNVL VTSHQAFTR	300
EAVTAIAATT LGNLDDFAAG RPCPHALKA	329
SEQ ID NO: 49	moltype = AA length = 396
FEATURE	Location/Qualifiers
source	1..396
	mol_type = protein
	organism = Escherichia coli
SEQUENCE: 49	
MIISAASDYR AAAQRILPPF LFHYMDGGAY SEYTLRRRN DLSEVALRQR ILKNMSDSL	60
ETTFLFNEKLS MPVALAPVGL CGMYARRGEV QAAKAADAHG IPFTLSTVSV CPIEEVAPAI	120

-continued

KRPMWFQLYV LRDRGFMRNA LERAKAAGCS TLVFTVDMPT PGARYRDAHS GMSGPNAAMR RYLQAVTHPQ WAWDVGVLNGE PHDLGNISAY LGKPTGLEDY IGWLGNNFDP SISWKDLEWI RDFWDGPMVI KGILDPEADAR DAVRGADGI VVSNNHGRQL DGVLSSARAL PAIADAVKGD IAILADSGIR NGLDVVRMIA LGADTVLLGR AFLYALATAG QAGVANLLNL IEKEMKVAMT LTGAKSISEI TQDSLVQGLG KELPAALAPM AKGNAA	180 240 300 360 396
 SEQ ID NO: 50	moltype = AA length = 320
FEATURE	Location/Qualifiers
source	1..320
	mol_type = protein
	organism = Pedicoccus sp
SEQUENCE: 50	
MSKIQNHQKV VLVDGAVGS SYAFAMAQQG IAEFFVIVDV VKDRTVGDAL DLEDATPFTA PKNIYSGEYS DCKDADLVV TAGAKPKPGE TRLDLVNKLN NILSTIVKPV VDSGFDGIFL VAANPVDILT YATWKFSGFP KEKVIGSGIS LDSARLVAL GKFKNVSPDS VDAYIMGEHG DSEFAAYSTA SIGTKPLLDI AKEEGVSTDDE LABIEDSVRN KAYEIIINKKG ATFYVGTLA MRISKAILRD ENAVLPVGAY MDGEYGLNDI YIGTPAVING QGLNRVIESP LSDDEMKKMT DSATTLKVKL TDGLKALEEK	60 120 180 240 300 320
 SEQ ID NO: 51	moltype = AA length = 316
FEATURE	Location/Qualifiers
source	1..316
	mol_type = protein
	organism = Acidobacteriia bacterium
SEQUENCE: 51	
MKAGVVGCCGF VGSTAAYTMA LGGVASELVL VDLVPALAKA HAEDILHATP FAAPRVVVG DYPQLEGAHV VILACVGQK PGETRLQLE RNAKVFQVV PQVLKAAPSA ILLVASNPVD IITRMVTHIS GLSPARVIGS GTILDTRARF TLLSEYLGVA PQSVHAYVLG EHGDSEVLAW SSAKVGGISL EIFAAQRGVS IDDTVRARID DGVRRAAYRI IEGKKATYFG IGAGLTRIVE AIRDDERVVL TVSICQDGLE GLEEVGLSLP RVVGRPGVIA TLHPPLSPEE QDALCHSARV LQQAASELGF LERARA	60 120 180 240 300 316
 SEQ ID NO: 52	moltype = AA length = 320
FEATURE	Location/Qualifiers
source	1..320
	mol_type = protein
	organism = Bifidobacterium longum subsp. longum
SEQUENCE: 52	
MAETTVKPTK LAVIGAGAVG STLAFAAAQR GIAREIVLED IAKERVEAEV LDMQHGSFY PTVSIDGSDD PEICRDADMV VITAGPRQKP GQSRLELVGA TVNLIKAIMP NLVKVAPNAI YMLITNPVDI ATHVAQKLTG LPENQIFGSG TNLDSARLRF LIAQQTGVNV KNVHAYIAGE HGDSEVPLWE SATIGGVPMC DWTPLPGHDW LDADKREEIH QEVKNAAYKI INGKGATNYA IGMSGVDIIIE AVLHDTNRIL PVSSMLKDFH GISDICMSVP TLLNRQGVNN TINTPVSDE LAALKRSAET LKETAQQFGF	60 120 180 240 300 320
 SEQ ID NO: 53	moltype = AA length = 334
FEATURE	Location/Qualifiers
source	1..334
	mol_type = protein
	organism = Helobdella robusta
SEQUENCE: 53	
MTSTTCVKEL FTHIIDEERH TSVKTVVGT GQVGMASAFQA MLTQGVISEL ALVDMVADKL KGEMMLDLQHG QAFLRTVKVQ ASTDYSVTAG SKICVVTAGA RQNLGETRLQ LVQKNVAIFK HIIPNLIKYS PNCILIIVSN PVDILTYVAW KISGLPRRNQV IGSGTMLDSS RFRFILSERI GIAPKSVHGY IIGEHGDSSV AVVSSVNIAG TRLKDISHPT GDPDDPENWN EIHKEVNNSA YEIIKLKGYT SWAIGVMIST LCNAILKNQK VIYSLSTLAK GYHGIEEEVF LSLPCVVGEK GVGAVFDQKL LPNEMEKVKA SAKTLDHVIK SLEL	60 120 180 240 300 334
 SEQ ID NO: 54	moltype = AA length = 316
FEATURE	Location/Qualifiers
source	1..316
	mol_type = protein
	organism = Plasmodium berghei
SEQUENCE: 54	
MAPKAKIVL GSGMMIGGVMA TLIVQKNLG VVMFDIVKNM PHGKALDTSH TNVMAYSNCK VSGSNTYDDL KDADVVIVTA GFTKAPGKSD KEWNRDDLLP LNNKIMIEIG GHIKNNCPNA FIIVVTNPVD VMVQLLHQHS GVPKKNKIVGL GGVLDTSRLK YYISQKLNVC PRDVNAHIVG AHGNKMVLK RYITVGGIPL QEFINNNKIT DQELDAIPDR TINTALEIVN LHASPYVAPA AAIEMAESY IRDLRKVLIC STLLEGQYGH KDIFAGTPLV IGGNGVEQVI ELQLNADEKK KFDEAVAETS RMKALI	60 120 180 240 300 316
 SEQ ID NO: 55	moltype = AA length = 308
FEATURE	Location/Qualifiers
source	1..308
	mol_type = protein
	organism = Armatimonadetes bacterium

-continued

SEQUENCE: 55
 MKVGIIGGG RVGSNAAYAL QLMGAVSEVV LLDVNREAAE GEALDLRHGA ALSQPIRFRA 60
 GDYADLEGAE MVIITAGLRR RPDESRLELI NRNVGLFRDI LSQLRQVSLH PNVIFLVSN 120
 PVDILTYLAV RESGLPAERV IGLGTVLDT C RFRSYLAEEF NVAATVDAL ILGEHGDSMV 180
 PIWSCATIAG VPLKDYPGYS KEAMEQIFQQ TRNAGAEAIR LKGAGYAVG VSITQVVNAV 240
 ALNQNAILPV STLQTGALGI QDVCFSLPTR VGRNGVGAI EIGVSNEERD ALQNSAKVLK 300
 ETLQKVIG 308

SEQ ID NO: 56 moltype = AA length = 250
 FEATURE Location/Qualifiers
 source 1..250
 mol_type = protein
 organism = Spodoptera frugiperda

SEQUENCE: 56
 YSITADSQC VIAAGVRQME GEDRRNLVQR NVEVLKGIVP KLLKYSPNAI LLIASNPDV 60
 LTYVTWKISG LPKHRVIGSG TNLDSARFRY LLSQKLNIAPI TSCHAYIIGE HGDSSPVWS 120
 SVNIAGVRLR DLNKNIGLDS DEENWKETHT QVQQSAYEVI KLKGYTSWAI GLSLTQLVKA 180
 ILTNAASVHA VTTCVEGEHG ITDNVFLSLP CVLGRNGVLD VIRQPLTDVE REQVFKSATL 240
 MAELQAKIKF 250

SEQ ID NO: 57 moltype = AA length = 318
 FEATURE Location/Qualifiers
 source 1..318
 mol_type = protein
 organism = Lachancea dasiensis

SEQUENCE: 57
 MSEVQPATTP VKVVIIGVGS VGSATAYTLL LSKIVSEIVL IDVNKDKAEG ESMQLNHAAS 60
 STTKSRAGDF PDCAGAAIVI VTCGINQKNG QTRMDLAAKN AAIMQEIVPN VAQHAPDSII 120
 LIATNPVDVL TYVSYKASGF PSSRVFGSGT VLDTARFKYI LGEHFQIPSE SVDACVIGEH 180
 GDGVPVWSL TNIDGMKLRE YCEARNHNVD EEVFHKIFEQ TRNAAYDIK RKGYTSYGINA 240
 AGLLRIVKAI LGDTNAVALTV STVGDYYGV E DIAISVPTRL DKDGVHPASE ISLNQKELGL 300
 MKKSADQIKS ALVALTQK 318

SEQ ID NO: 58 moltype = AA length = 327
 FEATURE Location/Qualifiers
 source 1..327
 mol_type = protein
 organism = Atribacter laminatus

SEQUENCE: 58
 MNQKKQKISL VGVGRVGSTL ALTILMKGLA QELVLVGENN HRAQGEAYDL IHASSFAHPI 60
 EIYAGPIEAT SHSDIIIVSA SVPMTNMKNR LDLGVGNAL FRNLIPQLSQ LSPEAIFI 120
 TNPVDIRMYY TIRLSNFPPS RVMTGTTLID SGFRRSLLGL YSGIHSEDIH AYILGEHGDS 180
 EFPVLSLADF CGGMGMEHSLP VCRKTCPLKN LTEVFEEAKD GGMLVYRHKG YTNAIALAT 240
 STLIEAIFND SGRILPVSTL CLSMPVVLGR EGILKIIELN MNEKEQAAFR 300
 SSAAVLKAT QEMGIEKNMS EQRKPKTE 327

SEQ ID NO: 59 moltype = AA length = 317
 FEATURE Location/Qualifiers
 source 1..317
 mol_type = protein
 organism = Bacteroidetes bacterium

SEQUENCE: 59
 MTSNQGKRKV GVVGTGLVGS SLVYALMIRE LATEIVLVDI DREKATGEMM DFNHGLFSK 60
 PVKITAGNYA DLKEAQVVI AAGASQKPG E TRLDLLARNV NIFRTIVPEV VQHNPTGIIIL 120
 IAATNPVDILT HISLKESQLT AGKVIGSGTI LDTSRFRPLL QOHYEVDARS VHAYIVGEHG 180
 DSEIPVWSLA NIGGVRLQEF APLKKNKRYDQ IEMDNLFAGV RDAAYEIIKR KGATYYAIGL 240
 GLVSIVETIL CNYRSVLSVS TLMTGQYGV S DMCLSLPCVV GANGIEEILN LNLSPGEEKG 300
 FRLSAEKLKA TLQSLGG 317

SEQ ID NO: 60 moltype = AA length = 328
 FEATURE Location/Qualifiers
 source 1..328
 mol_type = protein
 organism = Rhodohalobacter barkolensis

SEQUENCE: 60
 MKKQEDRNHQ KGSSAWKSRA VAIVGAGAVG ATPAYAMAQN GAADEICLID LNNELCKGQV 60
 LDLSHGLPFY PTINIIYAGSE KDYADADVIV ITAGAAQKKG ETRLDLLKKN SAIIEGIVDQ 120
 ITAQDSKATI VVVTNPVDIL TKIALERSGW DRSRVIGSGT VLDSSRFKYL LSEFFPNVYVG 180
 SIHAYILGEH GDSEFAAWSM ANISGVQLDE YSKQLGIDNW PEKKKEIELE VRDSAYHIID 240
 YKGATNFGVG LALVQIVGAI LKNQRRVLTV SYHLEGEYGI SDICLATPCL ISQNGVASII 300
 GTELTKEEEQE KLVNSANILK REYSELKR 328

SEQ ID NO: 61 moltype = AA length = 287
 FEATURE Location/Qualifiers
 source 1..287
 mol_type = protein

-continued

```

SEQUENCE: 61
organism = Piloderma croceum
LLLRPIASDI LLVDTDDLRL RAQVQDLSDA AFLSNTKIRP STPVEAGQCN IIVITAGAKS 60
RDGKSRRNLI DRNYLVLDNV IREMSPIRKD AVLLLVSNPV DVLAYFAQKM SGLPHAQVIG 120
SGTFLDSVRL RSALADQVQV ADTAVHAYVL GEHGDSQMMA WSTATVGGSP VQNFLSSTCP 180
SISDLANSAK TKAYSIIAAK GATAYGIASI VSSICESIIL NQRHVRPVSH WIESLETTYIS 240
LPAVLGHGGV QKTIEVPLDE TEKEELLQKSV EEIRSNIRT VEEGEQYQ 287

SEQ ID NO: 62      moltype = AA length = 316
FEATURE          Location/Qualifiers
source           1..316
mol_type = protein
organism = Caldithrix abyssi

SEQUENCE: 62
MKIGIIGCGY VGSTVAYAVL FKKAATEGRIVL IDKNHQRAVA EAADISHAIP FSHAIIDIYAG 60
DYSDLKGAEI VIIAAGVNQK PGETRQLQLLG RNEAVLRLQIL PPLLNEAPEA MILMTTNPVD 120
VMTHLAAKLA EELGYNGHHLV FGSGTTLDTA RFRSLLSAKL GVAPQHVHGY VIGEHGDSEV 180
LTWSIVDIGG LQLDDFVAAQ KLTFTDADKE QIDEQVRRAA YVIIEGKGAT YYGIGAAVAN 240
IVDVVAHQQK AILTICKPLK EVEGIADVTL SMPHLIHGKK RMMPLPLKLS QEERKALKQS 300
ASIICKFLDD FKNGAS 316

SEQ ID NO: 63      moltype = AA length = 287
FEATURE          Location/Qualifiers
source           1..287
mol_type = protein
organism = Candidatus Abawacabacteria bacterium

SEQUENCE: 63
MANVCKTEKI AIIAGGRVGT TFAFAALIRN IVAEVMLVDI DQERCKGEVL DL AHLGLSHVN 60
SGKVKYQGTLK QAATADIIVI TAGVAQKPGQ TRLQLVETNA KIIKGILKEM GKINPMAILI 120
IVSNPVDVLT YVAQKCSNL AGQIFGSGTD LD TARLQYLL GKHFNNHPRS VNAFIIGEHG 180
DSELAVWSSA NVAGIPIKKT RGYSPA KMRA YM TETRSAAG KIIKLKGATF FAISMALIEI 240
CTAILKDQNL VLPVSTYIEN YCGIKDVCLA VP AVIGRSQI KKLIKLP 287

SEQ ID NO: 64      moltype = AA length = 313
FEATURE          Location/Qualifiers
source           1..313
mol_type = protein
organism = Candidatus Abyssobacteri a bacterium

SEQUENCE: 64
MKQAKIVIVG AGAVGSTFAF TMMRSGLVGE IVL LDANRDR AE GEAMDLNH GLFFVPHVE 60
RVGEYSDCSG ASIVVITAGA KQRPGETRLD LVQRNTTICK EIMRQVMEYT RDSIILMVTN 120
PVDILTVQAY RISGLPSGVK IGSGT VLD SA RLFMLSRHC NVDSRN VHAY ILGEHGDSEV 180
A AWSLTHIGG IPILHYCSIC KIDCGSEQRK RIADA VRDSA YHVIESKGAT NFAVSLALEN 240
IVAAIVRDSN SVLT VSVPLK GEFGV SDVAM SVPVIVNREG VHHVLEPPLA EDELRALQRS 300
ASVLQDILRQ ITI 313

SEQ ID NO: 65      moltype = AA length = 310
FEATURE          Location/Qualifiers
source           1..310
mol_type = protein
organism = Candidatus Aenigmarchaeota archaeon

SEQUENCE: 65
MRKVKSIIGA GFVGSTAAYS LLISGACSEI VLIDVNQKA KGEAMDLKHG MQFVHQSKIR 60
FGEDYRLCKG SSIVIICAGA GQKPGETRLD LVKKNAAIFK DMIPKITKHC KDCILLVSN 120
PVDVLTIVIY EYSGFP PERV FGT GTVLD TA RLRYHLGEHY NVSASSVHAY ILGEHGDSEF 180
PVWSSARIGG APLKSMKEYS KNDMDRIAKE TKNAAYEII TKGATYYAIG LVITKIVKAV 240
FSDSNEVMVP STRLDGYYGI RDVCLSVPCV VGSNGIDKQL LIPLNATEKR SLRKSANILR 300
STIKEVMVKK 310

SEQ ID NO: 66      moltype = AA length = 308
FEATURE          Location/Qualifiers
source           1..308
mol_type = protein
organism = Candidatus Altiaarchaeales archaeon

SEQUENCE: 66
MRVGIVGAGN VGSTIAYTLS MGSSVSEI VL ADKNV DKA HGV EVL DLRHGLP FIP YT KLEYG 60
GLESLSNLDV VVL TAGIPRT HGESR L DLAG KNV GLF KEII PQLAG QNKNA ILL VVS NPVD 120
IMTYVALKYS GFPKNRVFGS GNVLDSARFR SML GH HF KVD PANVHSY ILG EHGD SAF PFF 180
SQAFIGCNPL KNMEGYD ERE VRE IFDRVKA VAAQVI KLKG ATYYAVS LGV N KIIESI DLD 240
KKRVM PVSTL IDR FYSGSEL CLS VP AVV GK NGIE KILH VP F NEE EKT LFQ AS VE KIA KTI 300
EEVGLSGG 308

SEQ ID NO: 67      moltype = AA length = 310
FEATURE          Location/Qualifiers
source           1..310
mol_type = protein

```

-continued

```

SEQUENCE: 67          organism = Candidatus Amesbacteria bacterium
MKSTFKVSIV GVGKVGATAA YAMCLNGRAT DLVLVAHQQE TAAEAKLDLE HALPFLQHIN 60
IVATDSYEAV SGSDLIVITA GVAQKPGQTR LDLLQQLNQL FTEIIPKIYA ASPESIVLII 120
TNPVNDILTYH SSQALPKPG QLFGSGTLLD TARFRFHLSE IFNVNPRSIH AYVLGEHGDH 180
SFPPALSAHHI SGQPLSQFPN YSLEKVDQAF TOTQQAAYKI ITAKGATYYA IATAILKLTN 240
TIYSDAKSIL PVSVPITNLY GVSGMALSMP CVVGQAGAEQ VISIPLSNEE QAQFIQAAADA 300
LKEAYRNVKK                                         310

SEQ ID NO: 68          moltype = AA length = 318
FEATURE
source
1..318
mol_type = protein
organism = Candidatus Aureabacteria bacterium

SEQUENCE: 68          moltype = AA length = 318
MKTVDVWINK VVIIGTGFVG SSIAFALMMS GRASEIVLID ANRKKAEGEM MDLNHGSAFT 60
PPVDIYVGDY KDCKNADIIIV IAAGANQKPG ETRIDLVHNK TAIFKEIIPQ ITAVEKNAIM 120
LVVTNPVDIL TYVTLKVSRL LPEQVIGSGT VLDTARFRYFL LSEHCGVDPY HVHAYIIGEH 180
GDTEVPVWSR VIIGVTPLEH FCKLCHKKCN EEVKNAIFHD VKNAAYQIE RKGATYYAVA 240
LGVLRIIESI LRNESSLTI SSYLDNYYGI NDVCLSPVSL LRSEGIQKHI ELPLNDEEHA 300
ALQRSADEMK NIIAQLDI                                         318

SEQ ID NO: 69          moltype = AA length = 307
FEATURE
source
1..307
mol_type = protein
organism = Candidatus Buchananbacteria bacterium

SEQUENCE: 69          moltype = AA length = 307
MSTKNKVTII GAGMVGSTAA YSLIGDDSTQ EIAMVDVNKR LVASQVMQLQ HSPVPGSFTK 60
VKVGSYSDIK DSKVVVITAG AAQKPGETR DLVKKNSAI GEIAPHIFKQ NPKAIVVVVS 120
NPVDVLTYQI IKMFPAKKNQ IIIGSGTLLD ARFRFLIGQK LDVNPRSVHA YIVGEHGDS 180
LPLWSTAMIG NTFLDKFHKL SDREKKQIFD AAKNAAYAII AGKQATYYAI ASGITSIVRA 240
ILFDQKTVLP ISHLLEGEGYR IRNICLSMPA IVGAAGVIKK VNPEISAQEK KQLQKSAQL 300
KTVTKGL                                         307

SEQ ID NO: 70          moltype = AA length = 306
FEATURE
source
1..306
mol_type = protein
organism = Candidatus Cerribacteria bacterium

SEQUENCE: 70          moltype = AA length = 306
MRCDCTRKSII IGLGNVGATA AYALLLEGVV DELVLYSRK EKAIGEKLLD EHGLNFLQPA 60
KITATDSFDD VANSDVIVIT AGVAQKPGET RLDLVKKNTT IIEALIPLL KAAPEAIYVI 120
VSNPVDLLTY KANILALPK GRVFGSGTML DTARFRYHLS EMLSINPRSI HAYILGEHGD 180
SSFPVLSSA VGGQPLTTFP SYSQDQALQA FESARTAAYN IIQAKGATYY AIGVVIVKLV 240
RTILQNSHSV LPISIPVEDY YQQYNVSLSV PCIVGRNGVE QILKAQLNEK EQEQFAKSAT 300
LLKSFL                                         306

SEQ ID NO: 71          moltype = AA length = 310
FEATURE
source
1..310
mol_type = protein
organism = Candidatus Collierbacteria bacterium

SEQUENCE: 71          moltype = AA length = 310
MLKSETFKVA IVGLGRVGMAT TAYALLLKGL CTELVLFSRE LAKAEGEKDD LTHGAPFYPH 60
ATVTATDKFE DLNGTDLVIF TAGCAQQPGQ TRLDLTKQNC EIVAKLIPKI VREAPQTLIL 120
MVANPLDDMT LRATELAHLA PGRVFGSGTLL DTARFRRLHL SKAISLNPOS IHAYVLGEHG 180
DSSFATVTTA TIGGKPLLSF PEMSPDHIAW SLAETRKDAG KIISSKGATY YGIATAVSHI 240
VETIMRNSRK VLPLSTVLTG QYGLTNIALS LPCILGRSGV ERVVDLPLSL EETASLHASA 300
QILKTHLDEL                                         310

SEQ ID NO: 72          moltype = AA length = 317
FEATURE
source
1..317
mol_type = protein
organism = Candidatus Falkowbacteria bacterium

SEQUENCE: 72          moltype = AA length = 317
MFKKKSPKPV ESTRVVIIGA GLVGATSAWA IMLQGIASEI VLIDIDKNKC AGEVLDLQHG 60
ISFLSPAKWV AGSYACDRDA DVIVITAGLK QKEGQARLEL AAANSKIVAE IMHKIIKYTS 120
EAIILMVTNP LDVLTIVAYR ASGFPANQVF GTGTALDSSR FRYLLAQSLG VAAESVGAYL 180
IGEHGDSSVP VYSHANIMGG KLAVFPKYNK QAVQAAYRQT KNAAYDLICK KGATYYAIAL 240
AVARIVRAIL YNENHIFPAS VYLSGQYGIS DVYLSLPAVV NRRGVKEILN IELSALEKKQ 300
LHQSAKVIRA AIDSIIK                                         317

SEQ ID NO: 73          moltype = AA length = 319
FEATURE
Location/Qualifiers

```

-continued

```

source          1..319
               mol_type = protein
               organism = Candidatus Firestonebacteria bacterium

SEQUENCE: 73
MTKNLLPKI SIIGAGNVGM RYAYALAIISG LARQINILDI NKEKAEGEVL DLSHGAPYY 60
PVEIVYGEYD SLANSDLIVI TTGKNQKPEQ TRLELVKENV ALYKNIIPAI MKYSPSSII 120
VVTNPVDVLA YTAFKLSGKP SKEVISSGTV LDSARLRFL SKHFEIDAHN VHAYILGEHG 180
DSEFPAYSSA TVGGIHICKDY CLKCGDKSCN YQDELARIFK EVKSSAYGII KRKGETSYGI 240
GLALVRITSA IIKDENSVLP VSNLINGYLG IKDVYLSIPA IVNKGGVREV LDIKLDSEE 300
GQLINSANVL KNIIKDSGM 319

SEQ ID NO: 74      moltype = AA length = 301
FEATURE
source          1..301
               mol_type = protein
               organism = Candidatus Gracilibacteria bacterium

SEQUENCE: 74
MSKKIVLIGT GMIGMSFIYS ALQSDLASEY ILIDIFTEYA EGNKLVEDI LPYNTKNISI 60
KVGDYSDCSD ANIIVICAGR PQKQGETRLE MIQDNAAIKI SIGVKVKDSG FSGISIIASN 120
PVDIRLTFIYQ QEYNFPKEKV IGSGCILDSDN RLKVEIAKEL NIFPHEVEAY VLGEHGDSSV 180
MQFSNIKING KYDDFSEEK KVEIGNKVRN KAYKIIAEAKK ATYFGIGVAL KKLCEDIIFD 240
KKEIVSVGHF PEKDNKKLYI SSPCIIGKKG IEEILNLDS KKEQEEFKNS VGKIYKALKE 300
Y 301

SEQ ID NO: 75      moltype = AA length = 322
FEATURE
source          1..322
               mol_type = protein
               organism = Heimdallarchaeota archaeon

SEQUENCE: 75
MRKLSPKVAI IGGGNVGIRY AYALMIKGLA RHIIIVVDIDI QKVQGEIMDL SHGIPYTPSV 60
KLEVGA耶EDQ QGSDLVVITA GKNRKPGE SR LELTKANVAL YKEIIPQIMK YAPSALFLVV 120
TNPDVILSYT AYKISNKPEV EVIGAGTVLD TARFRNLIAA HCKVDPRNHV GYIILGEHGET 180
EFPPVWSRLV GGNNLIKDYCP ICENCDICNH EEELDKIFIE VRDSGAKIIK KKRETSYGIG 240
LALVRITEAI PNDENAILPV SCLVNGFLDI HDVYLSLPAI ITKNGVKKVL EIKLNEQEQL 300
LIKSANKLK EKIKEIMALL EK 322

SEQ ID NO: 76      moltype = AA length = 328
FEATURE
source          1..328
               mol_type = protein
               organism = Candidatus Hydrogenedentes bacterium

SEQUENCE: 76
MDFSELNPKR KVVGAGSV GATFCYALAQ SGLADEIVLI DHNKDLARGQ VLDLAHGQSF 60
FPPVTIRVGD QDQDYADATLL VMTAGARQOP GETRLQLLRR NASIVRQVAR DIALANSPGT 120
LIVVSNPVDI LTKVAAETG WRRGRVVLGSG TVLDSARFRS LLSHNCIGIDI HNIHAYILGE 180
HGDSEFAAWS MTHVAGIPID EYCTSCRMCT PTGKQDEWLQ VRKSIEEQVR NSAYHIIDYK 240
GSTYFAIGLA LVRIAGAILR NENSVLTST LLEGEYGVSG LCLSVPAVS GRGVERIMEA 300
RLNPEEQKAL ARSAAVLRS LNELEGNE 328

SEQ ID NO: 77      moltype = AA length = 306
FEATURE
source          1..306
               mol_type = protein
               organism = Candidatus Jacksonbacteria bacterium

SEQUENCE: 77
MNKKITIIGA GMVGSTTAYS LLASNNVGEI ALIDINOKLA HAQAMDLQHS VPFLGYTSIH 60
NGTYADIKNS SIVIITCGAA QKPGETRLDL VKKNSAIKID ILIKVYHQNP HVIIIMVTNP 120
VDILTHIACT LYKPNKSCII GTGTILDSAR FRFLLGEYHL INPQSVHAYI VGEHGDSEVP 180
LWSTATVGNT PIEKVKLMSA TVKKEIFENA KNAAYAIIEG KQATYYAIAT GVAMLVDVIT 240
HKDKTVLPVS RYINGEYGIR NVCLSMMPAVI GKEGIISTIP LTLSQEKSA LKKSAKTLQS 300
ILKATD 306

SEQ ID NO: 78      moltype = AA length = 306
FEATURE
source          1..306
               mol_type = protein
               organism = Candidatus Kerfeldbacteria bacterium

SEQUENCE: 78
MSNNNGKVVVI GAGFVGATSA YAMFIDGAPS EIVL LDVNKE KAEGEAMDLE QGMQFVPGTK 60
LSYGSYDALV RNAEVVVITA GARTKPGQTR LELISVNAQI LKTIIQNIKK YNKDCCLLIV 120
TNPVDVLTLYL AVRYSGFSYD KVFGSGTTLD SARLYRYFLQG KLRVHPSSIH AYMLGEHGD 180
EFPWAWSARV CGVPVLEMEG LDRAILNRLA DNTRRAAYEI VARKGATYYA IGLVVAQLVR 240
AILDSSNTIF PLSVYLUKYH GVSDVTLSVP VVLRNRSGAHI RFQLPLSQQE KKQFRSCAKI 300
VRSLSQS 306

```

-continued

```

SEQ ID NO: 79      moltype = AA length = 308
FEATURE          Location/Qualifiers
source           1..308
mol_type = protein
organism = Candidatus Komeilibacteria bacterium

SEQUENCE: 79
MLLKNKVTTII GAGMVGSTAA YSLIASNLLE ELALIDSHPA YLKAQVMDLQ HAIPFHGNTQ 60
VKAGTYEDVR DSRVVVITCG AKQKVGETRL DLVKKNSAII KEILPVRVFAK NPrAVVIMVT 120
NPVDVLVTLA VSLYPKCARQ IMGSGTILDS ARFRFLLSQK LQVSPKSIHA YIVGEHGDS 180
LPLWSNATIG GGGIEKFAGL SAAAKKSLFN QAKNAAYAI AGKQSTYFAI ASGVAEIVGA 240
VIMNRRTVLP VSHLLTGEFG LKKVCLSLPV VVGREGIAQK VGLTISPAEK KQLVESARKL 300
KQFERSVK             308

SEQ ID NO: 80      moltype = AA length = 321
FEATURE          Location/Qualifiers
source           1..321
mol_type = protein
organism = Candidatus Latescibacteria bacterium

SEQUENCE: 80
MQDTSGSRK VVVVGAGAVG ATFAYALAQS GLADEIVLIID RNEDLVRQGV LDLAHQGPFF 60
PTVSIRIGDA SDYSSASLIV ITAGAAQRPG ETRLQLLQKN AAVIKGVMTD IVEQASPAAV 120
LVVTNPVDM TYVALETSGW DRSPRIGSGT VLDSARLEHL LSAHCGVDVH NVHGFVLGEH 180
GDTEFVAWSL THIAGMOMDV YCPICKGCAD NLAEREAITQ QVRDSAYHII GYKGATWFAV 240
GLALVRIAGA ILRNQRSQLT VSALLHGEFG LSDVCIGVPC IVSARGIERV VESALSESEL 300
AALHSSARTL RKAVEQLHAG A             321

SEQ ID NO: 81      moltype = AA length = 304
FEATURE          Location/Qualifiers
source           1..304
mol_type = protein
organism = Candidatus Levybacteria bacterium

SEQUENCE: 81
MGSKVTTIIGG GHVGTSAAFM MLLRETAREI VLLDRDKNMK MGEQLDFQHS LAFLATTKIV 60
AAQSYEDTKD SDVIIFTSGV AQVSGQSRLD LLKSNAEILE ATLPGVVKLS PESVVIIVSN 120
PVDALTYKAS RLLNLPKGRV FGTFTSLDSA RPRFYLSELL SINPKNIHAY VLGEHGDS 180
PVISTADVGG QSLINMKGIS KEQIPECFIK TRDAAKTIIA TKGATYYAIG VVISQLTHAV 240
LEDAKRIFPV SIPLDGEYQG KGVSLSVPCE LGRKGMEKIL EIPLSDEEKR LMENSANILK 300
KFIN             304

SEQ ID NO: 82      moltype = AA length = 316
FEATURE          Location/Qualifiers
source           1..316
mol_type = protein
organism = Candidatus Liptonbacteria bacterium

SEQUENCE: 82
MASKPNKTPI ESTRVVIIGS GLVGATSAAYA LMLQGSASEI VLVDIDKKRC AGEVLDLQHG 60
SSFLPEVKWI AGDFSDCKEA DVVVIITAGLK QKENQSRLEL AKTNAKIVSD ITKNITKYTK 120
KAVILVVTAWL LDIMTYTAWR ASKLPKNQVF GTGTVLDSSR FRHLLAEKIK IDPESMGAF 180
LGEGHDSSVP VFSHANVMGE PISSLKKYKG INTEDAYKVV RNAAYELISK KGATYYAIGL 240
VVARLVRAIL YDENHVFPVS VYADGHYDLK DIYSLPAVV GRNGIKQIE IKLSDQEKNL 300
LQKSAAIIKQ ALKELE             316

SEQ ID NO: 83      moltype = AA length = 306
FEATURE          Location/Qualifiers
source           1..306
mol_type = protein
organism = Lokiarchaeum sp.

SEQUENCE: 83
GNVGMRYAYA LMIKGIAKRL VLVLDLKERL YGEVMDLSHS TPFTSPVEVI SGDYSDIRDS 60
ELVVITAGRK QKPGQSRLDM AKGNVSLYEN MIPEVVRKHAP SAIYLIVSNP VDILTYAAYK 120
FSKPPASEVII GSGTVLDSSR FKYLGGKHCN VDPRNVHAYI LGEGHDTEVP IWSKAMLGG 180
ALIDYCPTCK NTMTCDREEE LNTIFEEVRD SAYEIIERKG ETSYGIGLSL VRITQAILND 240
ENSILPVSCF IDDYIGIQDV YLSLPAVNVK EGIRDVLKLE LDKGEQEKLRF SAYTLKNVL 300
KEVGLN             306

SEQ ID NO: 84      moltype = AA length = 309
FEATURE          Location/Qualifiers
source           1..309
mol_type = protein
organism = Candidatus Magasanikbacteria bacterium

SEQUENCE: 84
MDNKKPLATK IAIIGAGSVG STIAYVATLK NLAAEISLID INVKKEEGEV MDIADGMCFV 60
ETGCVKGTDF ASARDADIIV YTAGAPQAPG ETRLDLATKN RNILKSVFKK IARFKNTAIV 120
IIIVANPVDAL TAEAIRLTQI PRGRVFGTGT ALDTARLRTE LSHLFKISAQ NVHGFVLGEH 180
GNSEFVAWSS VTIGGIPTAK LKLLTAHART EIETRVRQEA YEIIISRKNTF FYGIGLITAD 240
IIIEAVLFQK KILPISILVE NWNDVDNIVL GVPAILGRAG VEKIWPLILS KQEQQKLTAS 300

```

-continued

AAMLKKYEV	309
SEQ ID NO: 85	moltype = AA length = 280
FEATURE	Location/Qualifiers
source	1..280
	mol_type = protein
	organism = <i>Candidatus Gastranaerophilales bacterium</i>
SEQUENCE: 85	
MNDIKETVNPF NKAVIVGCGF VGSASAFSLM QSGLFTELVL IDSDSKRAEG EALDISHGIP	60
FSKPMKIHAG DYDDVKNAGI VIITAGANQQ PGETRLLDVK KNINIFKTTII PEIKKRNFNG	120
ILLIVANPVD ILTYAAVAKLS GLSENVRIGS GTVLDTARLKK YAIGRHLNVD SRSVHAFIIG	180
EHDGDSIAWV SSANVGVPV NNFCEMRGHY DHONATEKIA TDVKNSAYEI ISMKKATNYG	240
IAMSVKRICE AIVRNERSIL PVSTIMHGEY GISDVSLSP	280
SEQ ID NO: 86	moltype = AA length = 307
FEATURE	Location/Qualifiers
source	1..307
	mol_type = protein
	organism = <i>Candidatus Microgenomates bacterium</i>
SEQUENCE: 86	
MFHQETFKVA IIGLGRVGIT TAYAILLKG ATELVLFSRE IEKAEGEKAD LEHGMPPFYPK	60
CNIASSNNYE DLSGTDLVIF TAGCAQKPGTR ELVQDKNNC AIVDALIPNI VKHAPHTLIL	120
MVANPVDMT LRAVEKALP EGRVFGSGTL LTDARFRLHLL SKGVKISAKS IHTYILGEHG	180
DSSFAATSTA TIGGKPILSF PEMSPEHIQW AQDQTRKD DAY KIINGKGATY YGIASAVTHI	240
VETIMRDSKK IIPLSTPLTG QYGLDNVSLS LPCILGRNGV ERIIDLPLSV TETAALHNSA	300
KMIRQSI	307
SEQ ID NO: 87	moltype = AA length = 309
FEATURE	Location/Qualifiers
source	1..309
	mol_type = protein
	organism = <i>Candidatus Moranbacteria bacterium</i>
SEQUENCE: 87	
MILMKNKVII GAGMVGSTVA YSLILQDIDV EIALIDLNEV LVKSQVMMDLE HSIPFVGRTD	60
VKVGDYDDCA DSAVAVITCG AAQKPGETRL DLVQKNAAI KKVIPAFDK NPDIILVMVT	120
NPVVDVLTHVA VSLFPEKKDQ IMGTGTILDS ARFRHLICKK LDVDPKSIHA YILGEHGDS	180
FPVWSTASIG NMQIGTCDCI SPEEQENIFT QARGAAYTII EGKQSTYYAI GAGCAHLIRA	240
IVRNKKSVLP VSHLIEGYD INDVCLSMPA IVGVTGLMGK LCIKLNKTEQ KQLKKTAEVL	300
KETFAKIN	309
SEQ ID NO: 88	moltype = AA length = 310
FEATURE	Location/Qualifiers
source	1..310
	mol_type = protein
	organism = <i>Candidatus Nomurabacteria bacterium</i>
SEQUENCE: 88	
MANKSSKVVI IGAGFVGSTA AYALLIDGTA SEIALIDVDR EKAEGEALDL KHSLQFTSQ	60
KITFGEDYNL CKDAIIVIC AGAHQKPGET RLDLAQKNAK IFKEMIPKIT RHNRDCILLI	120
VANPLLDILTY LAIKYSGFPS HRVFGSGTIL DTARFRYLLG EYFEVDITTSV HAYILGEHGD	180
SEFPVWSTAN IAGINLRLFD NYNKKKMGEI FKKTRNAAYE IIAKKGATYY AIGLGITKIV	240
KAILKNHNEV LPVSCLLKNY HGISDVCLCSV PAIINKDGIK KQLKLPLNPQ EKANLRKSAN	300
VLKNIIKKVV	310
SEQ ID NO: 89	moltype = AA length = 318
FEATURE	Location/Qualifiers
source	1..318
	mol_type = protein
	organism = <i>Candidatus Omnitrophica bacterium</i>
SEQUENCE: 89	
MENLRPKVSI IGCNVGMRY AYALIIRGLV RQIVIVDIDR QRLEGEMVMDL SHGAPYISPV	60
EIIAGDYPDI KNSDLVVITI GRKQKPGQSR TDLARDNVEL YRKVIPFIVR YAPKSIIILVV	120
TNPVDVLAYA AYKISGKPAK EVISSLGTVLD SSRFRFLLSK HCNIIDARNVH AYILGEHGDS	180
EFAVWSKAMI GGILFKDYCR ICKTNNTCSS KEELNKIPLF VRDSAYKIIE KKGETSYGIG	240
LTLVMMITEAI INDENAILPV SSLVDGYLGI IDVYLSLPAI LNREGVREVL EIELSPQEEK	300
ALINSANSIK KVIKEVGL	318
SEQ ID NO: 90	moltype = AA length = 319
FEATURE	Location/Qualifiers
source	1..319
	mol_type = protein
	organism = <i>Candidatus Pacebacteria bacterium</i>
SEQUENCE: 90	
MENKNKYVNC KAREDALGYK ISIIGCGKVG MTTAYSILHD GVVNDLLVD RNKSKIIGEQ	60
LDLEHGLSFL HHAHIDATDD YANIKGSDLV IITAGVAQRP GDTRLDDANK NLKIIGEVIP	120
IVQHSPESI ILIVSNPVDV LTYKAYQLAG LPKGRIFGSG TTLDTARFRF HLSEFLKVNP	180
RSIHAYILGE HGDNNSPPVIS SASVGGQPLA TIPGFSEERA EKAYSKARDA AYKIIASKGA	240

-continued

```

TYYAIGAVVA HISRMVLTD KSLPVSIP LGYKDISGIA LSVPCIVGRD GVEKILDVKM 300
SWDEKKLQN AAEVLLKLL 319

SEQ ID NO: 91      moltype = AA length = 315
FEATURE          Location/Qualifiers
source           1..315
mol_type = protein
organism = Candidatus Peregrinibacteria bacterium

SEQUENCE: 91
MLPCGHTAKI GIIGAGFGVT TFAFAAMIKN LAARIVLVDA NQECKCEGEVM DLAHGLNFVE 60
TSQIMHGNYQ DLADADAVVV CAGRQRPGE TRLDLLQDNV KILKSILEEF KQLHNQTCIL 120
VIVANPMDAL TYVAKKLSGL PENQVFGSGT VLDRSRLRFL LGEYFEVSSQ SMKAYVFGEH 180
GDSEFPVWSH ANIGGKPLQD FARYDAQKME EIFQQTAKNA YEIIKRKNAT YYGIALALTD 240
IVESILYDQN KVLPVSHTLA DYYGTKDVL SLPCVVCRRG IKETLDIPLN EEEVQKLHSS 300
ATTLRQSLDS IQNLL 315

SEQ ID NO: 92      moltype = AA length = 308
FEATURE          Location/Qualifiers
source           1..308
mol_type = protein
organism = Candidatus Roizmanbacteria bacterium

SEQUENCE: 92
MKSNKVVIIG GGRVGMSAAF AMLLKEKVGE IVLYYDRDIEK MKGEQLEFEH SLSFLGSTKI 60
TVASSNNDFE NSDVIVFTAG VAQPKGESRL ELVQKNTIEV KDILPAAVSQ SPNAVLMVT 120
NPVDVLTYEA SKILNLPKGR IFGTTGSLDT SRFRFHLSI LKINTKNIHA YILGEHGDS 180
FPVVSNAADVQ QQQLLSPVGI NEKIIINDCYI KVKEAAQAIQ QSKGATYYAI AVVSELVEA 240
VLSDSKRVYP VSVPLDGQYG LKDVLSVPC VIGRNGVEEI LEIDLSSIEK QNLDKSSEIL 300
KEFNGKKT 308

SEQ ID NO: 93      moltype = AA length = 308
FEATURE          Location/Qualifiers
source           1..308
mol_type = protein
organism = Candidatus Saccharibacteria bacterium

SEQUENCE: 93
MNKQKLVIVG AGGMVGATAA YACALRSVVE EIVLIDRDPD LAWGQAADIT DGMGIDRCVV 60
VRPGSYDIK TDDIVVITAG APQQPGQTRL ELLGVNAEIM RGTVRNIMRN GARPYILVVS 120
NPVDALTYVA LKESGLPKSR VFGTGTTLDT SRLKSYIADQ LDVHSREVDA YILGEHGDS 180
FATIESAQVG EVPLADYPGF KPAMVDGIEB QIRQRAYRVI ETKRSTYYAI GFVISKIVSA 240
LRSSSRVYP VCSLVEGEYG LHDVVLGLPS TICADGVKIL TGYPLNEREQ AALRHSAKVV 300
TEAIRSLE 308

SEQ ID NO: 94      moltype = AA length = 322
FEATURE          Location/Qualifiers
source           1..322
mol_type = protein
organism = Candidate division BRC1 bacterium

SEQUENCE: 94
MERHSWAKRR KVAVVGAGAV GSTFCYALAQ SGLAEEIAVI DRNENLVRQQ VLSDLVHGQPY 60
FPTVSIKVDI ASDYADAQLI VITAGAAQRP GETRLQLLQK NAAIVRGIVD DIFARESAAT 120
IVIVSNPVV LTAVALQKRAK GARGRIIGSG TVLDSARFRY LLSQRCGVDI HNVHAYILGE 180
HGDSEFAAWS MTHVAGMPVD EYALQFRPSE DWAAERRQIE REVRDSAYHI IEYKGATCFA 240
IGMALVRIAG AILRAQRSVL TVSVALEGFF GLDDVCLSVL AVVSEGGMAK IVESALPDNE 300
REALFASAAV LKQAMSQDLR DE 322

SEQ ID NO: 95      moltype = AA length = 314
FEATURE          Location/Qualifiers
source           1..314
mol_type = protein
organism = Candidatus Woesearchaeota archaeon

SEQUENCE: 95
MEATMAKRTG KVAAIGAGFV GSTAAYALLI EGAASEITLI DRHLQKAEGE AMDLRHGLQF 60
RADSRVYGT SYSLCRDAEI IIVVCAAGHQE KGETRLLDKV KNAAMPREMI PSIVKHKNKDC 120
ILVVSNPVD ILARLALTYS GFPASRVRFGT GTMLDTARFR FLLGERFGVS PESVHAYILG 180
EHGDSSFPVW STANIAGVPL RDFRQYDKRA MEGIYQQTKN AAYEVIARKG ATYYAIGLAV 240
ARLWKAILSD QNMVFAALS LKGYHGNGI CLSVPCCVNR EGIREQIVMP LNAEEKKKLK 300
KSAGVMKGIL RGIN 314

SEQ ID NO: 96      moltype = AA length = 305
FEATURE          Location/Qualifiers
source           1..305
mol_type = protein
organism = Candidatus Yanofskybacteria bacterium

SEQUENCE: 96
MENTMSKIAV IGAGSGVATT AYTLVVMKLN AEVILVDINE AKEEGEVMDI NDALSFVETG 60
KIKGGSYEDA AMADVIIITA GLPQKTGEQN RLELVNKNKE IIRSIFGQLK PLNPSAVIVI 120

```

-continued

VSNPVDIMAY YAQELSGLPR TQVFGTGTSL DTARLRTEIG LALDVNPQSV SGFVLGEHGE SEFVAWSTVN VGGVPIREKL SQDRIDEIAQ KVDDAKNII DRKGATFYGI AAVVSDVVES ILLDQDKVLP ISTRLENHSG ISGVCLGMPA VIGRQGIKEI WNLELVGSEK EKLRSAAEI RQYIL	180 240 300 305
 SEQ ID NO: 97 FEATURE source SEQUENCE: 97	 moltype = AA length = 312 Location/Qualifiers 1..312 mol_type = protein organism = <i>Candidatus Anoxychlamydiales bacterium</i>
 MQPKQKKISI IGAGSVGSSY AFSLMMMSGIA REVVIVDINK KKAFFGEALDL SHVQSFISPV EISYGNYEDI KDSDIVVITA GLNQKVNRQSR LDLVKDNTKI FKEIIPKIAK NAPNSILIVV TNPVDVLSYV AYKISDFPKC NVIGSGTVLD SSRRLRSLLSN FYKIDARNIH ANLIGEHGNT EFPPVWSKANI SGIPFDKCNL IKYKKDELEK IFEEEVKNSAA KIEAKGSTC YGIAISLLKI TKSILRDENS ILPVSSFIQN YLDVNDVYLS VPALIGSGGV KGILKMDLSN DEINNNFLTA KTIKDVKQI GF	 60 120 180 240 300 312
 SEQ ID NO: 98 FEATURE source SEQUENCE: 98	 moltype = AA length = 305 Location/Qualifiers 1..305 mol_type = protein organism = <i>Chloroherpeton thalassium</i>
 MKVGIIGAGF VGATAAFAMA MRGSCSEIVI VDADNAKAKA QASDIEHAVP FSFAMTVRDG DFQDLKGAKV IIISAGVNQK PGETRLQLE RNANIFRDIV PKVVAIESNA VIVVATNPVD ILTSLTBQLA GLPEQQVMGS GTTLD TARFR ALIGNELGVD PQHVHAYVIG EHGDSEVF AW SSANVAGLSI PSFCKARQVR WNEEIQSQITA DNRVKAAYHI IEGKGATYYG IGAVLARISE ALIRNHRAVL TVSANIPEFG VALSLPRLVS GKGDGLIGV QTNDEERAAL ERSASVLREA LSAIS	 60 120 180 240 300 305
 SEQ ID NO: 99 FEATURE source SEQUENCE: 99	 moltype = AA length = 319 Location/Qualifiers 1..319 mol_type = protein organism = <i>Bellilinea caldifistulae</i>
 MTHVFKKPTR VAIVGAGNVG ATFAYSLLIS GLASEIVLID ANHAKAEGEA MDLNHAMPLG RPARIWAGSY EDTAGAVITV ITAGRSRQPS ETTRLDLVQRN TEIFKQIIPK IVERNPNPGIL LIATNPVDIL TYVAFKISGF RRNHVIGSGT VLTDTARFRYI LSQHFVDPPR SVHAIIIGEH GDSEVPVWSL ANIAGMRLPQ YCAVNELGCA DETLNTIPEQ TRDAAYHIIE RKGATYYGIA SGLVRICEAI IRDQGTVLSV SSYIDEPYYG IEDVCLSLPT IVLDLGVERI IRLDLNEEV IGLQKSANLL KNIISQLKI	 60 120 180 240 300 319
 SEQ ID NO: 100 FEATURE source SEQUENCE: 100	 moltype = AA length = 291 Location/Qualifiers 1..291 mol_type = protein organism = <i>Mantoniella antarctica</i>
 MATPACSSGV RAIGYAGATH RGSRDSAYPR ACVPASVRHV NREETLQLLS RSRRCKSACV VQATGSVSLR EDLFGSKEPL FTDCRSSKVS IVGSGQVGLA CAYALINQAT CRQIVLHDIA PKMDRLEAEV ADLHGAEFV DSLDVVATAD FADTADSIV IIPAGARQNE GESRLALVAR NVAIFEDIIP KIAAASPAK LLIITNPBCDI MTHVALRLSG FPSNRVIGSG TALDTSRFRS LLAHKLAVDT GSVHGMVLGE HGDSSVVWLW QVMVGGDFE PAYSYIYPE P	 60 120 180 240 300 291
 SEQ ID NO: 101 FEATURE source SEQUENCE: 101	 moltype = AA length = 332 Location/Qualifiers 1..332 mol_type = protein organism = <i>Homo sapiens</i>
 MATLKDIQIY NLLKEEQTPQ NKITVVGVGA VGMACAISIL MKDLADELAL VDVIEDKLKG EMMDLQHGSL FLRTPKIVSG KDYNVTANSK LVIITAGARQ QEGESRLNLV QRNVNIFKFI IPNVVKYSPN CKLLIVSNPV SGFPKNRVIG SCGNLDSARF RYLMGERLGV HPLSCHGWVL GEHGDDSSPVW WSGMNVAGVS LKTLHPDLGT DKDKEQWKEV HKQVVESAYE VIKLKGYSW AIGLSVADLA ESIMKNLRRV HPVSTMIGKL YGIKDDVFLS VPCILGQNGI SDLVKVTLTS EEEARLKSA DTLWGIQKEL QF	 60 120 180 240 300 332
 SEQ ID NO: 102 FEATURE source SEQUENCE: 102	 moltype = AA length = 307 Location/Qualifiers 1..307 mol_type = protein organism = <i>Spizellomyces sp. 'palustris'</i>
 MPAKIAIIGG AGSVGATAAY AILLRRCASE LLLVLDLTDK CAAQVLDLSD AAFLSDTKVR	 60

-continued

MGTPKEAGQC DIVVTAGAK QRPGETRVDL IGRNRAILSS VIESMCPFQ STILLLIANP	120
VDVLTYFAQK LSGLPOSQVL GSGTYLDSAR LRSELSSRLE VAETAVHAYV LGEHGDSQFV	180
AWSSAHVANT PLLSLPAMKG VDRKKLAEDV KNKAYKIIEA KGATYFGIGG VAASICECII	240
FDLRHIRPLS HFIPELNVCL SFPVVLGHSG VKQTLKPPLT DEEEELLKKS AKEMRTIIEK	300
YELAPNP	307
 SEQ ID NO: 103	
FEATURE	moltype = AA length = 332
source	Location/Qualifiers
	1..332
	mol_type = protein
	organism = Stylophora pistillata
SEQUENCE: 103	
MATVQEKLMT KVYEEERAAP SKVTVVGVGQ VGMACAYSIM QQGICRELAL IDMIEDKLKG	60
EMLDLQHCSR FVKNVSIHAS TDYAVANSN LCIVTAGSRQ KEGESRRDLV QRNVNIFKAI	120
IPQLVKHSPN TILMIVSNPV DILTYWAWKI SGLPHERVIG SGTNLDTARF HFLISEKLNI	180
APMNVLGHWII GEHGDAVSPV WSGVNIAGVN LKDLHPIGS DGDLEKWNEV HKKVVNSAYE	240
VIKLKGYTSW AIGLSVSTMQ QALLRNQKNV HAISTLAKGF HGIEDDVFLS LPCVLGSHGV	300
MVVIKQTLDE NETKQQLQTC RTMNELQQSL QF	332
 SEQ ID NO: 104	
FEATURE	moltype = AA length = 310
source	Location/Qualifiers
	1..310
	mol_type = protein
	organism = Crenarchaeota archaeon
SEQUENCE: 104	
MKISIIGSSR LGSTTAFELI NRELEASEIVL VDIIPNLPQG EAMDLNQMSA EKGKNIVVYG	60
SNDYKDIDKT SIVIVTAGVA RPKGMTRMDL LKTNAPIVRD VTKEIVKYAK DSIIIVVTNP	120
MDVMTYVALK TSNFERNRVI GMGGQLDSAR YREILSQMLN ISRSSIKALV IGEHGETMFP	180
IPRFSSIAGYK RVSELLDAEK IKEABEKTRK IAAEVIALKG ATVFAPISCI ATLVESIVKD	240
KKDMLPVSAY LQGEYGVKDV CIGVPAIIGR NGIEKIIELD LNEEEKARFG ASVNTIRKAI	300
EEILTLKIIIE	310
 SEQ ID NO: 105	
FEATURE	moltype = AA length = 331
source	Location/Qualifiers
	1..331
	mol_type = protein
	organism = Microcystis aeruginosa
SEQUENCE: 105	
MLDKIFTPNP YAEQPAPLRP RKGVIIVGQ VGMACAYSIM IQDCFDELIL QDIATDKVEG	60
EVMDLRLHGMF FIEPTDLKMG TVADVGQNAQ VVIITAGAAQ KEGETRLHLL ERNVAIFRRI	120
LEDVAVYCPS ALILVVSNPV DIMTYVTLKI TNPPPSRVIG SGTVLDSGRL RSLLSTQLHV	180
DARNVHAYII GEHGDSSELAV WSSANIGGAR LLEGDWQDLS AANQESLTIEI FLQVKNAAYE	240
IICKRKGYTSY AIGLATTDIV KAILRSQERI LTISTLLDGQ YGLKDVCASI PSVVNEKGVI	300
KTLNLALSPPR ETQQLHNSAK IMRDLIDQLE I	331
 SEQ ID NO: 106	
FEATURE	moltype = AA length = 310
source	Location/Qualifiers
	1..310
	mol_type = protein
	organism = Thermus caldophilus
SEQUENCE: 106	
MKVGIVGSM VGSATAYALA LLGVAREVVL VLDRKLAQA HAEDILHATP FAHPVVVRAG	60
SYGDLEGARA VVLAAGVAQR PGETRLQLLD RNAQVFAQVV PRVLEAPEA VLLVATNPVD	120
VMTQVAYRLS ALPPGRVVG SGTILD TARFR ALLAEHLRV A PQSVHAYVLG EHGDSEVLW	180
SSAQVGGVPL LEFAEARGRA LSPEDRARID EGVRRAAYRI IEGKGATYYG IGAGLARLVR	240
AILTDEKGVY TVSAFTPEVE GVLEVSLSLP RILGAGGVEG TVYPVSLSPPE REALRSAEI	300
LKEAAAFALGF	310
 SEQ ID NO: 107	
FEATURE	moltype = AA length = 318
source	Location/Qualifiers
	1..318
	mol_type = protein
	organism = Dictyoglomus thermophilum
SEQUENCE: 107	
MKKILIVGAG AVGTSFAYSL IHKGLVVEIV LYDIDEKKAK GEALDLAHHGI YFTKPVEVRA	60
GGGLEEAKDSD IVVITAGAKQ RPGETRLQLL DRNISIYKDL VPEIVKNGFK GIFLIVTNPV	120
DVLTQFYATF SGFPRNRVIG SGTVLDSRF AYLLSKHCDV DPRSVNAYVI GEHGTAVLA	180
WSLTHIGGIS ISEFCPVCGR RCFDEDVKA IIKEVRDSAY KIIEYKGATY YAIGLALVNI	240
VEAIVRDENR ILPVSTVHPQ IFEFKDVPLS LPSIVGRNGV KKVLRVKLTE EEEKELYNSA	300
KFIKEAIDSF SSSLSSR	318
 SEQ ID NO: 108	
FEATURE	moltype = AA length = 371
source	Location/Qualifiers
	1..371
	mol_type = protein
	organism = Strongylocentrotus purpuratus

-continued

SEQUENCE: 108
MTTITSKLMD MVDHLDSSS PTKVTIVGVG QVGMACAYSI MTQNIASEIA LVDVIADKLK 60
GEVYDMQHGQ AFVKGCSVKG DTDYKVTAGS RLCIITAGAR QREGESRLNL VQRNVKIFEG 120
IPVNLVRYSP NTVLLVVSNP VDILTYVAWK LSGLPSNRVII GSGTNLDTAR FRFLIGEKG 180
IAPSSVHGYI IGEHGDSVA WWSSTTVAGV SLQQLDPEIG TVKDPENMHQ VHQEVIDSAY 240
EIKLKGYTS WAIGLSCATL AAAVLRNQKG VYAVSTVAKG YHGIEHPVFL SLPCILGQDG 300
ITHVIKQTLN TKEQAQLQAS ANTLWDIATS LDIREPSKYT NMLMGIGHFV MWTFVPLVAL 360
FAGHRAGHGR G 371

SEQ ID NO: 109 moltype = AA length = 320
FEATURE Location/Qualifiers
source 1..320
 mol_type = protein
 organism = Endomicrobium proavitum

SEQUENCE: 109
MADKLSPKVS IIGCGNVGMR YAYSMIIKGA ARELVLVDYN RQKAEGEAMD LSHGAPFVSP 60
INIIYAGDYPD TANSSDLVVIT AGRGQKPGET RIDLIKGNAE ILKSVPQVV KYSPKAIILV 120
ASNPVDILSY ITYKISGKPA NEVIGSGTVL DSARFRFLIG KHCVNDRSRI HASIPGEHGD 180
TEFPMWSKAM IGGVLFKDYC KVCNKENCNAK QEDAKLNEIF EDVRDSAYEI IAKKGETSYG 240
IGLALTAKISK AILKDENSVL GVSSLDDNYH GVSGIYLSPV AVVNKGIRQ TLQVDFDMVE 300
LDSFINSAEQ VKVIKASGF 320

SEQ ID NO: 110 moltype = AA length = 316
FEATURE Location/Qualifiers
source 1..316
 mol_type = protein
 organism = Methanophagales archaeon

SEQUENCE: 110
MPVSRSKVAI IGTGNVGSTF AFALMISGLA REIVLIGRDK KRAEGECMDL NHGASFVPPV 60
TIIYAAGYEGC AGADVVVITA GARQOPGESR LDLAQQRNADI FKDIIPRIAE HAENAILLVV 120
SNPMDVLTYL TIKLSGFSPQ SVIGSGTVLD SSRFRFLISQ QCQVDPRNH AYIIIGEHGDT 180
ELPVLSHANI GGMLFPQYCA RCNRCRDYKE EFGRIFEEVK NAAYQVIEAK GATYYGISA 240
LIRIVAAILR DENSVLPIST LLHDYYGIED VCLGIPAVVN RSGVERVLLL ELSPTEQELL 300
RHSARTLKEV IRGINL 316

SEQ ID NO: 111 moltype = AA length = 317
FEATURE Location/Qualifiers
source 1..317
 mol_type = protein
 organism = Geobacillus stearothermophilus

SEQUENCE: 111
MKNNNGGARVV VIGAGFVGAS YVFALMNQGI ADEIVLIDAN ESKAIGDAMD FNHGKVFAPK 60
PVDIWHGDYD DCRDADLVVI CAGANQKPGE TRLDLVDKNI AIFRSIVESV MASGPQGLFL 120
VATNPVDILYT YATWKFSGLP HERVIGSGTI LDTARFRFLG GEYFSVAPQN VHAYIIGEHG 180
DTELPVWSQA YIGVMPIRKL VESKGEEAQK DLERIFVMVR DAAYQIIEKK GATYYGIAMG 240
LARVTRAILH NENAILTVSA YLDGLYGERD VYIGVPAVIN RNGIREVIEI ELNDDEKNRF 300
HHSAAATLKS LARAFT 317

SEQ ID NO: 112 moltype = AA length = 314
FEATURE Location/Qualifiers
source 1..314
 mol_type = protein
 organism = Cetobacterium ceti

SEQUENCE: 112
MKTRKVGIIG VGHVGSHCAL AMILOQGACDE LTLDVDIRDKQK AKSQAIDCMD TISFLPHRTI 60
IKDGEIIDDLLT MDIIIVISVG NLTADKNRLN ELKGSIEVMK SFVPQIVKNG FKGIFVVITN 120
PVDIVTVYVQ KLSGFPHNRV IGTGTGLDSA RLKVKVLSQT NVDPRSIQAF MLGEHGDQS 180
ANFSTGTINC KPILDYLKEN EDTIGKLDDL QLEHKVQAQGA WDIYAGKNCT EYGIGCTCS 240
LVONIFHHNER RVIACSAYLQ GEYDYEGIYI GVPAIIGKDG LETIIELPLD ERERTKFKKS 300
CEIMSSYIEM AKNY 314

SEQ ID NO: 113 moltype = AA length = 305
FEATURE Location/Qualifiers
source 1..305
 mol_type = protein
 organism = Gemmatimonadetes bacterium

SEQUENCE: 113
MTISVGVVGTL GWVGSSVAIS TLHGGFASEL LLADVRHDLA EGEAMDLAHL AAFYTASVR 60
AVPIDEMLHT DALVVAAGRG GKPNECSRLL LRDNAKILRE LGEKLRGYRG LVVVVTNPVD 120
VLTYVVAESS GLPCERVIGT GTMLDTARLR QVLGHELRVD PHSVHAQVVG EHGDSEVVLW 180
SSAHVGGTPL REWPGWSRER EQPIATEVRT AAYEIIKRKG ATNHAIGLTT AALLRSALRG 240
ERRVLTWSRV QSGVLGLRDV ALSLPTVVDA GGAVDVIPVK LDDAERHGLD RSADVLRQAI 300
ASLRE 305

SEQ ID NO: 114 moltype = AA length = 314
FEATURE Location/Qualifiers

-continued

```

source          1..314
               mol_type = protein
               organism = Ignavibacteriales bacterium

SEQUENCE: 114
MKVGIIGSGM VGATSAAYAIM LRKAASEIVL IDSNEKRAQA EAADIIHAAP FTQATVVYAG 60
NYSDLKGAKI VVIAAGAHQS PGETRLMLME KNASILNDII SKVAEVAPNA IFLIATNPVD 120
ITYISISIA KNYGIPATRI IGSGGTLLDTA RFRSLLGSHTI GVDPQNVHAY VIGEHGDSEV 180
LTWSNIDIGG VPLEDLIEYR KIAFSEKIKE EIDKGVRNAA YKIIIEKGST YYGIGGAAIK 240
LVDVINRDNR SVLTISTYEE DVEGIKNVAL SLPHLIGGEG DLGVLPIRLN MKEKLLLKKS 300
AEVIRSKIDE YEVK                                         314

SEQ ID NO: 115      moltype = AA  length = 321
FEATURE           Location/Qualifiers
source            1..321
               mol_type = protein
               organism = Lentisphaerae bacterium

SEQUENCE: 115
MPQPSVHRRK VVVVGAGAVG STFAYALAOQS GLAEELIALID MNTDLAKGQV LDLVHGQPFF 60
PTVAIHVGTP ADYADAHLLIV LTAGAAQRPG ETRLQLLQKN AAVVRAVVRD IAAQDSRAVL 120
LVVSNPVDVM TAVALHETGW ERHRIVGSGT VLDSARLRHL LSAHCGVDVH NVHAYVLGEH 180
GDSEFVAWSM AHIGGLPMRE FCPLCRHCED WQAERRRIEQ QVRDSAYHII GYKGATWFAV 240
GMALVRTAGA ILRGQHSVLT VSSRLDGFEG LRDVCLSVPC VVASNGIERL IESTLPSDEQ 300
AALEASAALKL REAIGQVIGP A                                         321

SEQ ID NO: 116      moltype = AA  length = 302
FEATURE           Location/Qualifiers
source            1..302
               mol_type = protein
               organism = Rhizopus microsporus

SEQUENCE: 116
MVSCHKVAIV GAGAVGASTA YALMFKNICS EILIVDINPE IVQAQVLDDA DAASVSHPTI 60
RGASTEEFAQQ ADIIVITAGA KQKEDEPRTQ LIERNYRVLK NIIGSMQPIR PDAIILLVAN 120
PVWILTHIAQ RLSGLEPNQV IGSGYTYLDTT RLCVHLGEIF DVNPQSQINAF VLGEHGDSQM 180
IAWEAATIGG QPLTSPEFQ ELDKQAISSRA ISGKAMEIIR LKGATFYGIG ACAADLVHTI 240
MLNKKSVHPV SYVVEKYGVT FSMPAKLGWK GVEKIYEIPL SEEETLLLE SVKALQEIER 300
LY                                         302

SEQ ID NO: 117      moltype = AA  length = 377
FEATURE           Location/Qualifiers
source            1..377
               mol_type = protein
               organism = Meloidogyne enterolobii

SEQUENCE: 117
MAENGIFVPV APIESTPQNQ VTIVVGQVG MACAYSILQQ NIATEICLTD VLADKLQGEM 60
MDLQHGLAFT HNTCIVNAST DYAKTAGPSKI CVITAGCRQR EGESRLSLIE RNVVIFKGII 120
PQLVRHSPNT VFLVVSNPVD ILTYVTWKLS GLPKERVFSS GTNLDSARFR FLLSERLNIS 180
PCNCCHAFIG EHGDSSVAWV SGVNAGVNL SAQDLTTGTS NSNAKNDGK LEEEIHKKV 240
QSAYEIIRLK GTYSWAIGLS VASIVQGVMR NSRNVFALT NIKGIHGFD DIFLSLPTVL 300
GSNGVNPIVR QNLTPKELEQ LRGSATQLL EIQKNFEALNN MGNNNDKYVVG NILNIQYFFP 360
LNSVVSLNFF YFYLTIN                                         377

SEQ ID NO: 118      moltype = AA  length = 312
FEATURE           Location/Qualifiers
source            1..312
               mol_type = protein
               organism = Nitrospirae bacterium

SEQUENCE: 118
MGHTHRVAAI GCGRVGATTA FALLHTGQVE ELVLIIDTDA RAEGEAMDLA HATALLAPAR 60
VWAGDYADAA SAAIVIVAAG RASASLDESR LQLVGDNVRI TRQIVGELAA VGFSGCLLMA 120
TNPDVVLQAQI AASASPLPAA RVIGSGTLLD TARLRLQLLAT ACALDPRSVH AHILGEHGDS 180
EIVAWSSAAI AGMPWAAFGA QAGVELAPEA ILDSVRGAAP AIIERKGATN FAIGAALAR 240
VSCILRDEHS VLMVSTRLAG EYGERDVYLS VPCVVGGGGV ERIIELDSP AEWAGLHRSA 300
EILRDVARRV AA                                         312

SEQ ID NO: 119      moltype = AA  length = 333
FEATURE           Location/Qualifiers
source            1..333
               mol_type = protein
               organism = Trichomonas vaginalis

SEQUENCE: 119
MSAAAHVLIT GAAGQIGYIL SHWIASGELY GDRQVYLHLL DIPPAMNRLT ALTMELEDCA 60
FPHLAGFVAT TDPKAASFKDI DCAFLVASM LKPGQVRADL ISSNSVIFKN TGEYLSKWAK 120
PSVKVLVIGN PDNTNCEIAM LHAKNLKPEN FSSLSMLDQN RAYYEVASKL GVDVKDVHD 180
IVWGNHGESM VADLTQATFT KEGKTQKVVD VLHDHYVFDT FFKKIGHRAW DILEHRGFTS 240
AASPDKAAIQ HMKAWLFGTA PGEVLSMGIP VPEGNPYGIK PGVVFSFPCN VDKEGKIHVV 300
EGFKVNDWLR EKLDFTEKDL FHEKEIALNH LAQ                                         333

```

-continued

```

SEQ ID NO: 120      moltype = AA length = 303
FEATURE
source
1..303
mol_type = protein
organism = Stieleria varia
SEQUENCE: 120
MKITIVGTGK VGSIAIFANV MNPMASELML VNRTREKAEG DALDLSHASA LRNSNMRISA 60
GDIPDSEGS  IIVFTSSVPY GDPNRKRSEL ASDNHALLKE WIPALAKFSP GAVLIIVSNP 120
VDALTYAAIE LSGFDPCRVI GTTLDLDSVR YRAMLSDTLK IHSDDIRAYI LGEHGDTQFA 180
ASSIATMGG E RFYPNDVSQQ IFQRTVGMGY QVSKLKGYTIN YGIALATMQI LDSIVYDLRH 240
TMLPSVLIDG PCEVTDVCLS LPAVIGREGV TRVIRPQLTA EEQDAFRRSA ESVRKTIISLM 300
RAV
303

SEQ ID NO: 121      moltype = AA length = 380
FEATURE
source
1..380
mol_type = protein
organism = Pusillimonas sp.
SEQUENCE: 121
MTCVEDFRLR AMRRVPRMFY DYADSGSWTE STYRANETDL ERLKFRQRVA VNIESRNLRT 60
SMIGEDVAMP VAIAPTGLT MQHADGEILG AKAAEAFGIP FTLSTMSICS IEDVAHTSR 120
PFWFQLYVMP DRDFIGRLID RARAAACSLA VLTLDLQVMG QRHKDIKNGL STPPKPTLAN 180
LNLLLTKEPRW CIDMLGTQRR GFGNIIGHAK GVSDVGSLSA WTASQFDPTL SWNDIEWIKK 240
RWGGKLILKG VMDAADARIA AQCGADALVV SNHGGRQLDG APSSISALPA VVQAAGK DIE 300
VFVDDGGIRSG QDILRAVALG ARGTMIGRAF LYSLGAMQP GVQRCLQMLA NELDVSMAFC 360
GHTDIRKVRD SILVNGDIFD 380

SEQ ID NO: 122      moltype = AA length = 354
FEATURE
source
1..354
mol_type = protein
organism = Cyanidioschyzon merolae
SEQUENCE: 122
MTSGIDEAAA ELQGSLFGDV SPAIDTPGAA LARSYPVRT VVGAGDVGVA CAYNILTRDI 60
CSELVLVDVL KDKLKGVQMD LHGGAFYST RIRAEESYED TAHSACVIT AGVRQPGES 120
RLELMDRNAA LFKGIIPPLV QYSPTNTLLV VSNPV DLLTH LAWQMSGLPR ERVIGSGTYL 180
DSSRFRTLLA QRLGIDTASV QAMVLGEHGD SSFVYRSGIT VGGVPLRTCF ERMTDAASAS 240
TAFYDLVKGV HQQVAAAAYE VIKLKGYTINW AIGSAVGSI V TTIVHDRRKV LPITTHAGSL 300
RGLESADVFL SLPCVLRNG VVEVLQILPF MESDEKEDLQ SSIEALQSTP KKAS 354

SEQ ID NO: 123      moltype = AA length = 320
FEATURE
source
1..320
mol_type = protein
organism = Spirochaeta africana
SEQUENCE: 123
MIKNKIAIIG AGSVGATIAY NLSNRGLANE MVLVDVNREK AEAEVLDITH GMPLGSTANI 60
YAAGYEACGD AEIAIITAGA KQRPDES RQVQ LMDRNVGIMR SMVGDLMASG FGGVILVVTN 120
PVDVLTTFVAY RESGLPAHQV IGSGTVDSA RLRTFLSRSC SINPQNVHGY VIGEHGDTSF 180
PAWSNVTFGG VGVQEPCQDC DAGCCDDELR RQATEYVRQA AQTIIKAKGS TYYAVAQAVA 240
ITQAVLRDE RRILPITAVM SGFEDFREVA FSYPHFVGKN GIATDAGING RLRYELTAEE 300
QEQLRTSIRY IIGNTEQAGY 320

SEQ ID NO: 124      moltype = AA length = 350
FEATURE
source
1..350
mol_type = protein
organism = Rosa chinensis
SEQUENCE: 124
MHKSTSASSL GPGGLDLTQA FFKPIHGAAL PSSSNGQT KI SVIGAGNVGM AIAQILTQD 60
LADELVLVDA IPDKLRLGEML DLQHAAAFLP RTKINASV DY AVTAGDSL CI VTAGARQIHG 120
ESRLNLLQRN VALFRKIIPP LAKYSPETIL LIVSNPV DVL TYVAWKL SGF PSNRVVGSGT 180
NLDSSRPFRL IADHLDVNAQ DVQAYIVGEH GDSSVALWSS ISVGGVPVLS FLKKQQIAYE 240
KETLESIHKA VVDSAYEVIS LKGYTWSAIG YSAAGLARSI LRDQRSIHPV SVLAKGFYGV 300
DGGDVFLSLP AQLGRSGV LG VTNVHLTDEE EQKLRD SAKT ILEVQSQLGL 350

SEQ ID NO: 125      moltype = AA length = 319
FEATURE
source
1..319
mol_type = protein
organism = Fretibacterium fastidiosum
SEQUENCE: 125
MQGKKRTVGI IGI GHVGAHV AYSLAVQGIA DEI ILVDRER KKADCEAQDV MDSV CYLPHR 60
VEVRSGDFPD LKDCDVLVNS AGHIQLLATG NKS RLEEMDF TIRAVNGYVD RVME SGFDGV 120
VINVTNP CDV VAWRFAELSG LPRGRVFGTG TGLDTARLRS ALAR QTGV DH KSICAY VMGE 180

```

-continued

HGASQMVPWS SVFGGRPLA	EWERTDERFR FDREAIRKEA	RDAGWVFSG KQCTEYGICS	240
TAARMASVVL LDEKQILPVT	TRLEGEYGES GFIGVPAVV	GAGGAEQVVE VPMSAELAE	300
FKGCCDDVRA NMAHALEVR			319

SEQ ID NO: 126	moltype = AA length = 363
FEATURE	Location/Qualifiers
source	1..363
	mol_type = protein
	organism = Hypsibius dujardini
SEQUENCE: 126	
MAPPAPPGAV PATAAPSPAT AAPSPAAAAP VGGDIRGKLM KDIQAPPETS LTKVTVVVG	60
QVGMAIAFSI VNRHLAGELC LVDVDEKKVK GELMDLQQGL QFAGNMKIQG STNYDITAGS	120
KLCIISAGAR QKDGEDRRAL LDRNIAIFKG IIPQLMKYSP DAVLCVVSNP VDVMTYLTWK	180
ISGLPPHRVF GSGCSDLSSR LRHMAERLK IDVSSCHGFV IGEHGESSVV VWSGLNVAGV	240
RLADIKPDIG TDKDSEKWVD VRHQVIDSAT EIIKLKGYTS WAIGLCVSEL CRAVIGNQRH	300
VFLSLSTMAKG YQGIDQEIVF SLPCLVLSQNG VNAAVINNLTLT ATEKEKLKKS ATELAEFLFKS	360
AKL	363

SEQ ID NO: 127	moltype = AA length = 321
FEATURE	Location/Qualifiers
source	1..321
	mol_type = protein
	organism = Mycoplasma wenyonii
SEQUENCE: 127	

METVKIAIVG CGAVGSSFLY SAIHQGLASE YGLLDYNTEF ARGQALDLED AIPHFDPRPR	60
IKVITDYKEL KNYQFIVITA GRAQKEGETR LQMIRDNAVIA MKGIAEQVKS SGFSGIVLIC	120
SNPVDVLYTI PQKVTFGDFEK KVIGSGTVLD TNRLKVEVIK GMDVAPASLE GAFVLGEHGD	180
SSLVTFSLMK VSGLSINRCE EVSYPNSPFC CSDYEELKEK VVYKKAYEII SRKRATHYGI	240
GVAMTKILKT IIYNTKEILP VSSILHGEYG LSDVAISVPT VVGSNGIERI AIIPLKEKEQ	300
NKLVASAKIL QENIAKVRDL I	321

SEQ ID NO: 128	moltype = AA length = 319
FEATURE	Location/Qualifiers
source	1..319
	mol_type = protein
	organism = Thermotoga maritima
SEQUENCE: 128	

MKIGIGVGLGR VGSSTAFALL MKGFAREMVL IDVDKKRAEG DALDLIHGTP FTRRANIYAG	60
DYADLKGSDV VIVAAGVPQK PGETRLQLLG RNRARVMKEIA RNRSKYAPDS IVIVVTPNPVD	120
VLYTFFLTKES GMDPRKVFGS GTVLDTARLR TLIAQHCGFS PRSVHVYVIG EHGDSEVPVW	180
SGAMIGGIPQ QNMCQICQKC DSKILENFAE KTKRAAYEII ERKGATHYAI ALAVADIVES	240
IFFDEKRVLT LSVYLEDYLG VKDLCISVPV TLGKHGVERI LELNLNNEEL EAFRKSASIL	300
KNAINETIAE ENKHQNTSG	319

SEQ ID NO: 129	moltype = AA length = 310
FEATURE	Location/Qualifiers
source	1..310
	mol_type = protein
	organism = Verrucomicrobia subdivision 3 bacterium
SEQUENCE: 129	

MKVGVIAGF VGSTGAYAMI LQGVASEVVL VDINRELAQA QAEDMLHAIP FSPSARVEAG	60
DYQDLRGARL VVLACGVGQK PGETRLQLLS RNAAVFRQVV PEVIKQAPEA VLMVSNPVD	120
LTHIVTKLA DLPAARIIGS GTILDSTARFR ALLAEEMAVT SHSVQAYVVG EHGDSEVLVW	180
SSAQIGGLPL EDFAAQVKRP VTAEIEERID KGVRQAAAYI IKGKAATYYG IGAGIARLAR	240
AIRDDEERAVL TVSNABEIKPA QLPPVALSLP RVVGARGVLS TLRPTLSFQE RDLLERSAHL	300
LLANVQELGF	310

SEQ ID NO: 130	moltype = AA length = 312
FEATURE	Location/Qualifiers
source	1..312
	mol_type = protein
	organism = Smittium mucronatum
SEQUENCE: 130	

MNFTHRVSII GGGGNVGAAT AFALVVMQLP VEILLLGASE KSARAQALDI RDAAHFSTA	60
CRAGTSEEVG CSDIIIVITAG ARQRGPGEPRS NLIDRNYVIL DSIIKKIQPI KPTAIILMVT	120
NPVDVLTSLA QKLSGLPKSQ VIGSGTFLDS GRLRNLYLSRK LNVNSNSIHA SMLGEHGDQ	180
FVGWSVASIA GSPLLKHPLM ANVDAEIEEN AIAHQAYEII EAKGSTYFGV GYHVAALIKC	240
ILTDGHKIYP VCNYNAKYDT YLSVPAILGS DGATPINLNL SDEEEVKLER CAESVNHSLG	300
QVNEFAVHVN EE	312

SEQ ID NO: 131	moltype = AA length = 531
FEATURE	Location/Qualifiers
source	1..531
	mol_type = protein
	organism = Methylobacillus flagellatus
SEQUENCE: 131	

-continued

MRDMFTKDGR RFEKFSLEAE GLLLDYSKHR ITDETLPLLF KLARDSKVEE WRDRMFAGEK	60
INFTEENRAVL HTALRNRSNT PVYVDGKDMV PDVNRVLAQM RKFSTSIIRSG EWKGYSGKRI	120
TDVVNIGIGG SDLGPVMVCG ALKPYAQAGL NAHFVSNIDG THLAQTLERC DPETTLFIVA	180
SKTFTTQETM TNARSARSWF LQAQKDEAHV AKHFVAISTN ADEVGKFGID AANMPEFWDW	240
VGGGRYSLWSA IGLSIAIYIG MDHFEELLQG GYEIDRHFKN APLEQNIPVI MALLGIWYNN	300
FFGADSLAIL PYDQGLARFP AYLQQADMES NGKFVARDGR IVQCTTGPII WGEAGTNQGH	360
AFYQLIHQGN RLIPCDFMMP LQSHYSMGKN GNEHHILLILA NCFAQTSALM QGKTLDDEAKA	420
ELVAQQLTGE ALETLLPHKV FEGNRPSTTI LFDKLTPKTL GKLIAIYEHK IFVQGIWNI	480
NSFDQWGVEY GKQIAKKILP QLSEDQASTE YDSSTNGLMN YFKSRTQSSG T	531
 SEQ ID NO: 132 moltype = AA length = 209	
FEATURE Location/Qualifiers	
source 1..209	
mol_type = protein	
organism = Methylobacillus flagellatus	
SEQUENCE: 132	
MSTLDDLANHG VPIPPIVINK VEDAVPMAEA LLEGGIKVLE VTLRTPVALQ AMEAIAKAVP	60
DAIVGAGTVR NIKAQNCND VGCQFAVSPG YTSELGRAAR EMGLPPLLPGV STGSEIMQAN	120
ADDYYFLKLF PAVAVGGINL LKGFFAGPFGD VKFCPTGGLT VESAPQFLAL PNVVVCGGTW	180
LTPGDAVARK DWAHITKLAR EASAIKAAA	209
 SEQ ID NO: 133 moltype = AA length = 226	
FEATURE Location/Qualifiers	
source 1..226	
mol_type = protein	
organism = Pseudomonas putida	
SEQUENCE: 133	
MTTLLERPOPK LSMADKAARI DAICEKARIL PVITIAREED ILPLADALAA GGIRTLEVTL	60
RSQHGLKAIQ VLREQRPELC VGAGTVLDRS MFAAVEAAGA QFVVTGPITE DILEAGVDSE	120
IPLLPGLISTP SEIMMGYALG YRRFKLFPAB ISCGVAAIKA FGGPFGDIRF CPTGGVPAN	180
VRNYMALPNV MCVGTGWMLD SSWIKNGDWA RIEACSAEAI ALLDAN	226
 SEQ ID NO: 134 moltype = AA length = 213	
FEATURE Location/Qualifiers	
source 1..213	
mol_type = protein	
organism = Escherichia coli	
SEQUENCE: 134	
MKNWKTSAA ILTTGPPVVPV IIVVKKLEHAV PMAKALVAGG VRVLEVTLRT ECAVDAIRAI	60
AKEVPEAIVG AGTVLNPOQL AEVTEAGAQF AISPGLEPL LKAATEGTIP LIPGISTVSE	120
LMLGMGYGLK EFKFFPAEAN GGVKALQATA GPFSQVRFCP TGGISPANYR DYLAALKSVLC	180
IGGSWLVPAD ALEAGDYDRI TKLAREAVEG AKL	213
 SEQ ID NO: 135 moltype = AA length = 228	
FEATURE Location/Qualifiers	
source 1..228	
mol_type = protein	
organism = Methylobacillus flagellatus	
SEQUENCE: 135	
MAKPLVQMAL DSDLDFDQTVA LATTVAPHVD ILEIGTPCIK YNGIKLLETI RAKFPNNKIL	60
VDLKTMDAGF YEAEPFYKAG ADIVTFLGTA DIGTIKGVID VANKYGKKAQ VDLINVTDKA	120
ARTKEVAKLG AHIIGVHTGL DQQAAGQTPF ADLNLVSSLN LGVDISVAGG VKATTAKQVV	180
DAGATIVVAG AAIYGAADPA AAAAEISAAA KGTQSSGGLF GWLKKLFS	228
 SEQ ID NO: 136 moltype = AA length = 181	
FEATURE Location/Qualifiers	
source 1..181	
mol_type = protein	
organism = Methylobacillus flagellatus	
SEQUENCE: 136	
MNKYQELVNV KLTNVINNTA EGYDDKILSM VDAAGRTFLG GAGRSSLVSR FFAMRLVHAG	60
YQVSMVGEVV TPSIQAGDLF IVISGSGSTE TLMPLVRKAK SQGAKVIVIS MKAQSPMAEL	120
ADLVVPIGGN DAHAFDKTHG MPMGTIFELS TLWFLEATIA KLIDQKGLTE EGMRAIHANL	180
E	181
 SEQ ID NO: 137 moltype = DNA length = 1002	
FEATURE Location/Qualifiers	
source 1..1002	
mol_type = genomic DNA	
organism = Methylobacillus flagellatus	
SEQUENCE: 137	
atgcgcatacg tcttttcag caccaggta tatgaccggg aaagtttct ggccctcccc	60
aggctccaga atgcccagtt catttccaa cagcccaac tcacgttga tactgccgtg	120
cttgccacag ggcggaaagt cgtctgtgcc ttcatcaatg acgacctgtc cgccgtctgt	180
ctggagagac tggcagccaa cggtaccaag ctgattgcac tgccgtccgc aggtacaac	240
catatcgacc tggaggctgc gcaaagactc aacctggcag tggcgcgcgt gccccctat	300

-continued

tcgccccatg	ccattgcgga	acataccgtc	gcattgatcc	tggccctcaa	tagacacctg	360
acgcgcgcct	acaaccgtac	ccgcggagggt	gatttcagcc	tgcggggctt	gacagggttc	420
gaccttagtcc	acaagacagt	aggcatcatc	ggcacccggac	agatcggtgc	agctttgcc	480
cgcattatgg	caggtttcgg	ctgcaggata	ctggcttacg	atcccacccc	caaccggaa	540
gtcattgcgc	tcggcgccag	ctacccatgg	ctcgatacgc	tgttggcgca	atcccacatc	600
gtcagcgtgc	attggccgtt	gagecctgco	acccaccatc	tcatcaatgc	cggttcgtc	660
agccgcgtc	agccaggcag	catgetcatc	aacaccaggc	cgggcgaact	cgtggatacg	720
cggccgtga	tccaaagcgt	qaagagcga	cacccatggc	atcttgggt	ggacgtctat	780
gaagaagaag	cagacctgtt	tttcgaggac	ctatccgact	ttccgttgc	ggatgacgta	840
ctggcgccgc	tgctgacctt	tcccaacgtc	atcatcaact	cccaccaggc	gttttcacac	900
cggaaagcgc	tcaatgccc	tgccagcacc	acggttagaca	atatacgcgc	ctgggcagca	960
ggccaccctc	agaaccttgtt	caggctggaa	acaggaagct	ga		1002

SEQ ID NO: 138 moltype = DNA length = 1002
 FEATURE Location/Qualifiers
 source 1..1002
 mol_type = genomic DNA
 organism = *Methylobacillus* sp

SEQUENCE: 138

atgcgcatcg	tcttttcag	cacccaggta	tatgaccggg	aaagtttct	ggccctcccc	60
aggctccaga	atgcccagtt	cattttccaa	cagcccaaac	tcacgttga	tactggcg	120
cttgcacag	ggcgaaagt	cgtctgtcc	ttcatcaatg	acgacgtgc	cggccccgt	180
ctggagagac	ttgcggccag	cggcacaag	ctgattgcac	tgcgtccgc	aggtacaac	240
catatcgacc	tggaggctgc	gcaaaactc	aacctggccg	tgttgcgtgt	gccccctat	300
tcgccccatg	ccattgcgga	acataccgtc	gcattgatcc	tggccctcaa	tagacacctg	360
acgcgcgcct	acaaccgtac	ccgcggagggt	gatttcagcc	tgcggggctt	gacagggttc	420
gaccttagtcc	acaagacagt	aggcatcatc	ggcacccggac	agatcggtgc	agctttgcc	480
cgcattatgg	caggtttcgg	ctgcaggata	ctggcttacg	atcccacccc	caaccggaa	540
gtcattgcgc	tcggcgccag	ctacccatgg	ctcgatacgc	tgttggcgca	atcccacatc	600
gttagcgtgc	attggccgtt	gagccctgco	acccaccatc	tcatcaatgc	cggttcgtc	660
agccgcgtc	agccaggcag	catgetcatc	aacccaggc	cgggcgaact	cgtggatacg	720
cggccgtga	tccaaagcgt	qaagagcga	cacccatggc	atcttgggt	ggacgtctat	780
gaagaagaag	cagacctgtt	tttcgaggac	ctatccgact	ttccgttgc	ggatgacgta	840
ctggcgccgc	tgctgacctt	tcccaacgtc	atcatcaact	cccaccaggc	gttttcacac	900
cggaaagcgc	tcaatgccc	tgccagcacc	acggttagaca	atatacgcgc	ctgggcagca	960
ggccaccctc	agaaccttgtt	caggctggaa	acaggaagct	ga		1002

SEQ ID NO: 139 moltype = DNA length = 996
 FEATURE Location/Qualifiers
 source 1..996
 mol_type = genomic DNA
 organism = *Methylobacillus rhizosphaerae*

SEQUENCE: 139

atgcgtatcg	tcttttcag	cacccaggta	tatgacaagg	aaagtttct	ggcttcccc	60
aggctccaga	acgctgatt	catcttcag	caacccaac	tgacgttgc	taccggcg	120
cttgcgcgtg	gagctgaatg	cgtctgtcc	ttcatcaatg	acgacgtgc	agcaccagt	180
ctggaaac	ttggccggccaa	cggcacaaa	ctgatcgac	tgcgtccagc	aggtacaac	240
catatcgacc	tggaaaggcc	gcaacgactt	aacctggctg	tgcgtccgt	accatctat	300
tcaccacatg	ccatggccg	gcataccgtc	gccttgcattc	tgcgtccaa	caggcattt	360
acgcgtgcct	acaaccgcac	gcfgaagggt	gatttcagcc	tgcgtccgg	gacaggtttc	420
gatctgtca	acaaggcgtt	aggegttgc	ggcacagggc	aaatcgccgc	aacgtttgc	480
cgcattatgg	caggtttcgg	ttcagatgg	ctggccat	atccgtatcc	caaccggca	540
gtcactgcgc	ttggggccag	ttacccatgg	ctcgatacc	tattggcgca	atcccattatc	600
gtcagcgtgc	attggccgtt	gagttccgaa	accatcac	tgtatcaatgc	cagctcg	660
agccgcgtc	agcaggccag	catgtatc	aacccaggc	gttgcgttac	gttggacacg	720
cctggccgtc	ttggaggatt	gaaacgtgg	cacttggct	accttggct	ggatgtgtat	780
gaagaagagg	cagatgttt	cttcgaggac	ctatccgatt	ttcccttgc	ggacgtatgt	840
ctggcacgc	tgctgacctt	ccccaaacgtc	atcatcaact	cgacaccaggc	tttttcacac	900
cggaaagcgc	ttgtatgcgt	tgccggccacc	acgcttgc	acatcgacgc	ttgggcggca	960
ggcaggccgc	aaaaccttgtt	caggctggaa	acatgta			996

SEQ ID NO: 140 moltype = DNA length = 996
 FEATURE Location/Qualifiers
 source 1..996
 mol_type = genomic DNA
 organism = *Methylobacillus methanolivorans*

SEQUENCE: 140

atgcgtatcg	tttttttcag	cacccaggcc	tatgacagg	agagcttct	cgcatttccc	60
aaggccccaga	atgcggattt	catttttcag	caacccaac	tcactacaga	taccggcg	120
cttgcgcgtg	gagcgaaagt	cgtctgtcc	ttcatcaatg	atgatttac	tgcgtccgt	180
ctagagaaat	tagctctgg	cggcacaaa	ctgattgcac	tgcgtccggc	aggtataat	240
catattgacc	ttggaaaggcc	gcaagaaacta	gggttggcag	tgcgtccgt	accatcgat	300
tcaccacatg	ccatggccg	acataccgt	gcgtatcc	tgcgtccgt	ccggcacct	360
acacgtgcct	acaaccgtac	gcfgaaggc	gacttcagcc	tgcgtccgg	aaccggctt	420
gatctgtca	acaagaccgt	aggegttgc	ggcacaggcc	agataggcgc	agcatttgc	480
cgcattatgg	cgggatttgg	ctgcaagg	ctacccatccc	caaccggag		540

-continued

gttatcgccg	taggagcatg	ctatctcgat	cttgatacat	tgcgtggaca	atcgcatatt	600
gttagcgtgc	actgcccgtc	gagcccaagg	acgcatacc	tgcataatgc	taactcactc	660
gccccatgc	ggcggtggcg	atgtgtgatc	aataccaggc	gccccgcatt	agtggataca	720
cctgcgtga	tttgcgtcc	gaaaagcgcc	catctagggt	acctggggct	agatgtttat	780
gaagaagaag	gggactgtt	cttcgaggat	ttatcgaggat	ttccgttgca	agacgacgt	840
ctggcgccgc	tgctgacc	ccccaaacgt	atcatcactc	cgccataggc	gttcttcaca	900
cgtgaagcg	ttggatccat	tgccagacc	accttgttca	atatcgagca	atgggcaacg	960
ggcacgcgc	aaaacctgg	caggctggaa	gcatga			996

SEQ ID NO: 141	moltype = DNA	length = 996				
FEATURE	Location/Qualifiers					
source	1..996					
	mol_type = genomic DNA					
	organism = <i>Methylobacillus arboreus</i>					
SEQUENCE: 141						
atcgcatcg	tcttttcag	tacccagggt	tatgatcg	agagcttct	cgcattaccc	60
aggctccaa	atacggaa	cgtattccag	caacccaagc	taacagcaga	tacagcggt	120
cttgctcg	ggcgagaa	tggtgtgc	tttatcaat	acgatgttc	tgccttgcgt	180
ttggaaacac	ttgggggg	ttgttacaaa	ctgatcg	ttcgctcg	cggttacaaac	240
catattatgc	ttgaggccgc	acacacact	aagctagaag	tgcgtccgt	gccagctt	300
tcgcccacat	ccattggcg	gcataccgt	gcttattatc	tgcactcaa	tcgcccac	360
acccgcgcct	acaacccgt	gcgcgaaggd	gatttca	tcgcgtgtt	cactggctt	420
gatctgttca	acaagacgt	aggtgttca	ggtacgggg	aaattgggac	ggcggttgc	480
aggatcatgg	ctgggttgc	ctgaagata	cttgcattac	atccctatcc	caacccggag	540
gttagaagccc	ttgggggtgc	tttactggaa	ctggatgag	tgcgtggc	atcgagatc	600
gtgagcgtc	actgcccgt	gagccggcc	acttaccat	tgcataacga	aagtgcgtt	660
aaccgcgtc	acgcgtggc	catgtatc	aataccaggc	gtggcgaact	gttagacacg	720
cctaccgt	tttgcgtt	aaagagcggg	cacctgggt	atctggcct	ggatgtctat	780
gaagagggaa	gggatttt	cttgcaggat	ttgtcg	ttccgttgc	ggacgatgt	840
cttgcagg	tactgacc	tcccaacgt	atcatcacc	cgccatcaag	ttttctcaca	900
catgaagcg	ttggatccat	tgctgacc	acccttgaca	acatcg	ctgggctgaa	960
gttagggcc	aaaacttgg	gaagctggaa	acatga			996

SEQ ID NO: 142	moltype = DNA	length = 996				
FEATURE	Location/Qualifiers					
source	1..996					
	mol_type = genomic DNA					
	organism = <i>Methylobacillus caricis</i>					
SEQUENCE: 142						
atgcgtatcg	tcttttcag	cacccagg	tatgaccgt	aaagcttct	cgagcacgac	60
aaacccaac	atgcagaatt	tgtattccag	caacccaac	tgactttga	aacggcggt	120
cttgcgcgt	gggctgagg	agtttgc	ttcatcaat	acgatctgc	cgctccgtt	180
cttgcgttca	ttgcacccaa	cgttacac	ttgtatcg	tcgcgttac	cggttacaa	240
catattatgc	tttgcgttca	cacacgtt	ggcttgc	tagcgttac	accatctat	300
tcgcccacat	ccattggcg	acatactgc	gccttgc	tgcactcaa	ccgcccac	360
acacgggtt	ataatcg	gagggagg	gacttgc	tcgcgttct	cacgggtt	420
gacttgc	aaacagacgt	gggtatc	gttacgggg	agataggagc	tgcatttg	480
aatatcgat	ttgggttgc	ctgcaaa	acttattac	atcttattac	caaccccta	540
gttagaaat	ttgggtgc	atatcg	ttgtatg	tttttggca	ggcccgat	600
gtcagcgtc	atgttcc	gagtc	acttattac	tgcataat	gagcacacta	660
ggccggatgc	actctggc	catgtatc	aataccaggc	gtggcgaact	gttggatact	720
cctgcgttca	tttgcgttca	gaaaaggc	cacctgggt	acctgggt	ggatgtctat	780
gaagaagagg	ccgatgtt	tttgcaggat	ttgtcgat	ttccgttgc	ggatgtgt	840
ctggcaccc	tgctgactt	ttctaacgt	atcatcact	cccatcagg	attctcag	900
cgcgaggccc	tgaatcgat	tgcaagcaca	acccttagcc	acatcagc	atgggcccag	960
ggatcaccgc	aaaacttgg	caacataga	ggatag			996

SEQ ID NO: 143	moltype = DNA	length = 996				
FEATURE	Location/Qualifiers					
source	1..996					
	mol_type = genomic DNA					
	organism = <i>Methylobacillus gramineus</i>					
SEQUENCE: 143						
atgcgtatcg	tcttttcag	cacccagact	tatgaccgt	aaagtttct	cggggttgat	60
aaggccac	gcccac	atgtattcc	caacccaa	tgaccgtc	tactcggtt	120
cttgcgcgt	ggcgacagg	agttgtgc	ttcatcaat	atgatctgc	cgccccat	180
cttgcgttca	ttgcaccc	aggcacac	ttgattgc	tcgcgttgc	cggttacaa	240
catatcgacc	tttgcgttca	acagcg	ggcttgc	tagtgcgt	accgttctat	300
tcgcccacat	ccattggcg	acataccatc	gcctgata	ttgcactcaa	tcggcac	360
acgcgttca	ataacccgg	gcttacac	ttgtatgc	tcagaggct	cacaggctt	420
gtatcgat	ggggattata	ttgtatgc	ttgtatgc	ttgtatgc	ttgtatgc	480
cggatcatgg	ccggatttgg	ctgcaaa	acttgc	tttgcgtat	acccttac	540
gtgaccgcac	tcgggtcg	atactgc	ttgtatgc	tattcagcc	ggcccgat	600
gtcagcgtc	actgttcc	aagcccaac	accttgc	tgcataat	tgagacact	660
aaccgcgtc	aaacgcgg	tatgtatc	aataccaggc	gtggtgc	ttgtatgc	720
ccagcgccgt	ttggggat	gaaaaggc	cacctgggt	atcttggct	ggatgtat	780

-continued

gaggaggagg cagacctgtt ttttggaggat cttccgatt ttcccttgca agatgtatgc	840
cttgcacgtc tgctgacttt tcccaacgtc atcatcacgg cacatcaggc tttccatcaca	900
cgtgaagcac tacatcgat tgcaagtacc accttggcca atatcagcgc ctggccgca	960
ggatccccat gaaatcttgtt aaaggcagaa gtttga	996
SEQ ID NO: 144 moltype = DNA length = 993	
FEATURE Location/Qualifiers	
source 1..993	
mol_type = genomic DNA	
organism = Methylobacillus glycogenes	
 SEQUENCE: 144	
atgcaaatcg ttttttcag caccaggcc tacgacccgtt aaagttttct ctcgcacccc	60
gcattggatg actgcgcagct ggtttccag caacccaaac tgactgttga taccggcgtt	120
ctcgcccgac gggccggatg ggtgtgtcg ttcataatcg atgatcttc cgcacgtgtt	180
ctggaaaaggc tggcccgagg cggcacacgt ttgattgcac tgccgtcgcc aggctataac	240
catattatcg tttaaaggccgc acaggccctg qccgtqccgg tggtgacagt gcccctatc	300
tgcggccacg cgattggccga gcacaccatc gcaactgtat tgccgttcaa ccgtcaccc	360
aaccggatc ataaaggccac ccgcgaaggd gatttcagcc tgcgegggtt gactggctt	420
gatctgttga acaagacggt gggcaggccaa aaataggcgc ggcttcgccc	480
cgcattatcg cgggtttgg ttcaggctgtt ctggctttagt acccttaccc caatccttgag	540
gtaatagcac tggggggccga gtatctggca ctggaaacagt tgtttccgca agcacatatc	600
gtcagcgttc actgtccgtt aagccggcc acgcattatc tgattaaacc ccagaccctg	660
gcgcgtatgc aaaaggcag catgtgtatc aataccagcc gggccgcctt ggtcgacacg	720
cccgccgttga ttgaggccat taaaactggc catctggccc atttgggtt ggatgtctat	780
gaagaagaag cagactgtt cttctctgtat ctttccgact atccgtgtca ggatgtatgtt	840
ttggccacacg tgggttgcacccatc atcatcacggc cgcatcaggc attttgacg	900
cggaggccg tggctggat tgccaggcc acgctgtgtt atgttgcgc ctggccggcgt	960
ggtcagccgc agaatcttgtt gcaggccaggc tag	993
SEQ ID NO: 145 moltype = DNA length = 990	
FEATURE Location/Qualifiers	
source 1..990	
mol_type = genomic DNA	
organism = Escherichia coli	
 SEQUENCE: 145	
atgaaactcg ccgttatag cacaaaacag tacgacaaga agtacctgca acagggttac	60
gagtcccttg gtttggatct ggaattttt gactttctgc tgacggaaaa aaccgctaaa	120
actgccaatg gtcgtcaagc ggttatgtatt ttctgttacg atgacggcag ccggccgtt	180
cttgcggatc tggaaaaggc cggcgatctaa tatatacgcc tgcgtgttgc cgggttcaat	240
aacgtcgacc ttgacccggc aaaagaactgtt gggctgttgc tagtccgttgc tccagctat	300
gatcccgagg ccgttgcgttgc acacggccatc ggtatgtatc tgacgtgttgc ccgggtt	360
caccggcggtt atcagcgatc ccgtgtatgtt aaccttcttc ttggaaagggtt gaccggctt	420
acttatgttgc gaaaacggcc aggcgttgcgtt ggttccggatc aaatcggttgc ggcgtatgttgc	480
cgcattatcg aagggtttgg ttcaggctgtt ctggcgatccatc aagtgcacgc	540
ggcgctggatc tccgtgttgc tgatgtgttgc ctggccaaacc ttgttcttgc atcagacgtt	600
atctctctgc actgcggcgct gacacggaa aactatcatc ttgttgcacgc agccgcctt	660
gaacatgtatc gatgtgttgc aataccatgc ggggttgcattt gattgtatct	720
caggcaggcaa ttggatcgatc gaaaatgttgc gtttgggtat ggacgtgtat	780
gagaacgaaac gcgatctt ctttggat aataccaaacg acgtgtatca ggatgtatgtt	840
ttccgtcgcc ttttgcgttgc ccacaacgtt ctgtttaccg ggcaccaggc attctgtaca	900
gcagaaggctc tgaccatgtt ttctcgatc acgctgttgc aacttaaaggca tctggaaaa	960
ggcgaaacctt gcccgttacg actgtgtttaa	990
SEQ ID NO: 146 moltype = DNA length = 1191	
FEATURE Location/Qualifiers	
source 1..1191	
mol_type = genomic DNA	
organism = Escherichia coli	
 SEQUENCE: 146	
atgattatcccgccatcg cgattatcgcc gcccgcggc aacgcatttc gccgcgttc	60
ctgttccact atatggatgg tgggtcatat tctgtatatac ctcgtccgcg caacgtgttgc	120
gattttgtatc aagtggcgct ggcgcacgtt attctgtatc acatgttgcg cttttttgt	180
gaaaacgcgcg ttttgcgttgc gaaattgtcg atggccgttgc cactggctcc ggtgggttt	240
tggccatgtt atgcgtgttgc tggcgatgtt cggcgatccaa aacggggcggc cgcgcgttgc	300
atcccggttgc ttttgcgttgc gtttgggttgc tgcccgatgtt aagaatgttgc cgcacccatc	360
aaacgcgcgc ttttgcgttgc gtttgggttgc ctggcgatccatc gcccgtttat gcttgcgttgc	420
ctggcgatgtt gaaaacgcgcg ggggttgcgtt acgctgttgc ttaccgtgttgc tatgttgcaca	480
ccggggcgac gtcaccgttgc tgccgttgc ggtatgtatc gcccgttgc ggcacatgcgc	540
cgttacttgc aacgcgttgc acatccgcgaa tggggcgatgttgc gacacgtgttgc	600
ccacatgttgc taggtatatac ctcgttgcgttgc ctggccaaac cgcacggactt ggaatgttgc	660
atccgtcgcc ttttgcgttgc ctcgttgcgttgc ttttgcgttgc ttaccgtgttgc tatgttgcaca	720
cgcgttgcgtt gggatggccc gatgttgcgtt acggttgcgtt gcaatgttgc gatgttgcgtt	780
gatgttgcgtt gttttgttgc tgatgttgcattt gtttttttgcgttgc accacgttgcgttgc	840
gacgggttgc ttttgcgttgc ctcgttgcgttgc ctttttttgcgttgc gaaatgttgc	900
atagccatcc ttttgcgttgc ctcgttgcgttgc ctttttttgcgttgc aacgggttgcgttgc	960
ctcggttgcgtt gtcaccgttgc gtcaccgttgc gtcaccgttgc aacacgttgcgttgc	1020

-continued

-continued

SEQ ID NO: 150	moltype = DNA length = 1596
FEATURE	Location/Qualifiers
source	1..1596
	mol_type = genomic DNA
	organism = Methylobacillus flagellatus
SEQUENCE: 150	
atgcgggaca ttttacaa ggatggcg agattcgaaa agttctccct ggaagcgaa	60
ggctgttc tggactactc caagcaccgc atcacggatc aaaccttgc gctactttc	120
aagctcgccg cgcaactccaa ggtcgaggaa tggctgtaccgc gcatgttgc gggcgagaag	180
atcaacttta cggaaacccg tgctgtacta cataccgcct tgcgaacccg tagaatacg	240
cccgatatacg tggatggcaaa ggatgtcatg cccgacgtca acccgtaact ggcccgatg	300
cgttaatcca gcacccatcg ctgcgttggg gaatggaaa gctactcagg caacgcatt	360
accgtatcg tcaacatcg categggcgc tcggacatcg gcccagtcat ggtgtgcggg	420
gcgcgtcaacg ctatgccca ggcgtggctt aacgcgcatt tgctgtccaa tatecgccgc	480
accacacccg cccagaccc tgaacgcgtc gatccggaaa ccacttttgc catgtcgct	540
tccaaagactt tcacccacca ggacccatcg accaatgcgc gctctgcgg cagctgttc	600
ctgcaggcg ccaaggatga ggcgcacgtt gcaaaacact tgcgtcgat ttccaccaat	660
gcggacaaag tccgcgaatg cggccatcg gcccacatcg tgctgtccatt ctggatgg	720
gtaggcattcg gctatccccct atggctgcgc atcggtctt cgatgtccat ctacatcgcc	780
atggatattt tcgaggatgt gtcgaaggd ggtacaaatcc tgcacccggca tttcaagaat	840
gcgcgcgtgg agcagaacat tcccgatcg atggcgatcg tgggcattcg gtacaacaaat	900
ttcttcgggtt cagacggccct cgccatctcg ctttcacgcg agggttgc cccgttccct	960
gcctacatcg acgaacggca tatggaaatg aacggcaatg ttgtcgcccg agacggccgc	1020
atttgtcaat gcacccatcg cccatcatcg tggggcggg gtcgcaccaa cggcccgat	1080
gcgttctacc aactgtatcc ccaaggcaac cgatgtatcg cctgcgttgcatt catgtgc	1140
ctgcagaccc actacacatcg gggaaaaat gcaatgaac accacccat cctgtgtggcc	1200
aattgttcg cggcaacccg cgcgtatcg cagggcaaga cactgtatcg agccaaaggcc	1260
gaacttgggtt cgcaaggatgt gactgtggaa ggcgtggaaa ccctgtgcgc gcacaaatgc	1320
tttgaggaaa accgcggcaag caccaccatcg ctgttcgacaa agctgacccc aaagaccctg	1380
ggcaagctga tgcgcgtatca cgaacacaaatg atattgtgcgc aaggcatcat ctggaaatc	1440
aactcccttg accaaatgggg cgttagataatc ggcacaaatcc ttgccaagaa aatctgcct	1500
caactgtatcg aagatcgccg atcgatgtatcg tacgacatcg cgcacaaatgg cctgtatgt	1560
tacttcacatcg cgcggacacaatc atccctggcc acctaa	1596
SEQ ID NO: 151	moltype = DNA length = 630
FEATURE	Location/Qualifiers
source	1..630
	mol_type = genomic DNA
	organism = Methylobacillus flagellatus
SEQUENCE: 151	
atgagcacac tcgatttagc caaccatggc ccagtcatcc cggtcattgt catcaacaaa	60
gtggaaagacg cgggtccgat ggcagaagct ttgtcgaaatg gtggatccaa ggtactggaa	120
gtgacccatgc gcaacgcgtt ggcgttcgtt ggcgttgcgg gcatcgccaa ggccgtgcct	180
gatgcctatcg tgggtggccgg gcatgttgcg aacatcaagg atgcacaaa ctgcacaaatg	240
gttggctgtc aatccgcgtt tagcccgatc tataccatgtt aactggcccg tgctgcgtt	300
gaaatggggat tgcattgttgc gcttgcgttgc tgcacccgtt gcgatgtatcg gcaacaaat	360
gcgcgttgcgtt actatccatcg ccaatgttgc cccgttgcgtt ctgtgggtgg cattaaccta	420
ctcaaggatcg tggccggccc gtcggcgat gtcacgttgc ggcacccggg cgggtcaac	480
gttgcgttgcgtt ctcccaatcg tctggcgatc cttacacgttgc tggtttcgcc cggcacatcg	540
ctcaacgcgtt gcgatgcgttgc ggcgcacaaatcg gattggcgcc atatcaacaaatcg gtcacccatcg	600
gaggccatcg tgcacatcg tgcacccatcg	630
SEQ ID NO: 152	moltype = DNA length = 681
FEATURE	Location/Qualifiers
source	1..681
	mol_type = genomic DNA
	organism = Pseudomonas putida
SEQUENCE: 152	
atgaccaccc tccaaacgcgc acacccaaatcg ctctcgatcg cgcacaaaggc cgccggatc	60
gtatgcgttgc gcaacccatcg ggcgttgcgtt ccatcgcccg tggtttcgcc ggtttcgcc	120
atccctgcgc tggccatcg cttggccgcgc ggcgttgcgtt ccatcgcccg tggtttcgcc	180
cgatgttgcgttgc atgggttgcgttgc ggcgttgcgtt ccatcgcccg tggtttcgcc	240
gttggggatcg gacccatcgatcg ggtatcgatcg atgttcgcgcg ctgtcgatcg tgcggccgc	300
cgttgcgttgc tccaccccgatcg cttggccatcg gacccatcgatcg aacccatcgatcg tgcggccgc	360
atccctgcgttgc tggccatcgatcg cttggccatcg gacccatcgatcg aacccatcgatcg tgcggccgc	420
tacccgcgttgc tcaaggatcgatcg cttggccatcg gacccatcgatcg aacccatcgatcg tgcggccgc	480
tttggccgcgc gtcgttgcgttgc tttggccgcgc gtcgttgcgttgc tttggccgcgc gtcgttgcgttgc	540
gtacgtatcgatcg acatggcattgc gcccaacatcgatcg atgttcgcgcg ctgtcgatcg tgcggccgc	600
acgacgttgcgttgc tcaaggatcgatcg cttggccatcg gacccatcgatcg aacccatcgatcg tgcggccgc	660
gcactgttgcgttgc tcaaggatcgatcg cttggccatcg gacccatcgatcg aacccatcgatcg tgcggccgc	681
SEQ ID NO: 153	moltype = DNA length = 642
FEATURE	Location/Qualifiers
source	1..642
	mol_type = genomic DNA
	organism = Escherichia coli

-continued

SEQUENCE: 153

```

atgaaaaact ggaaaacaag tgcagaatca atcctgacca cggcccggt tgcaccgtt 60
atcgtggtaa aaaaactggaa acacgcggg cggatggcaa aagcgttgg tgcgtgggg 120
gtgcgcgttc tggaaagtgc tctgcgtacc gagtgtgcgg ttgacgcata ccgtgcatac 180
gccaagaag tgcctgaagc gattgtgggt gcccgtacgg tgctgaatcc acagcagctg 240
gcagaagtca ctgaacgcggg tgcacagttt gcaatttagcc cgggtgtgg cgagccgtg 300
ctgaaagctg ctaccgaagg gactattctt ctgatccggg ggtacagcac tgcgtgggg 360
ctgatgtgg gtatgtacta cgggtggaaa gagttcaaat tttccggc tgaagctaac 420
ggccggcgtga aaggccgtca ggcgtacgg ggtccgttcc cccaggtccg tttctggcc 480
acgggtgtta ttctccggc taatcaccgt gactacccgt gctgtggaaag cgtgtgtgc 540
atcgggtgtt ctgggtgtt tccggcgtat ggcgtggaaag cggccgatata cgaccgcatt 600
actaaggcgtc cgcgtgaagc tgcgtggaa gctaaggctt aa 642

```

SEQ ID NO: 154 moltype = DNA length = 687
FEATURE Location/Qualifiers
source 1..687
mol_type = genomic DNA
organism = Methylobacillus flagellatus

SEQUENCE: 154

```

gtggcaaac cattagttca aatggcatta gattcactag atttcgatca gactgttagcg 60
cttgctacga ctgttgcacc acatgttcat attcttggaa tcgggtactcc ttgttatcaag 120
tacaacggta tcaagggtgtt ggagacttcg cggccaaatg tccctaaacaa caagatctg 180
gttgacatcg agaccatggg tgctgtttt tgcgtgggg agcatttttca caaggcagggt 240
ggccgatcg tgacccgtgtc cggcactgtt gacattttca cggatccagg cgtcattttt 300
gttgccaaaca aatacggcaaa gaaggctcaaa gtcgacccgttca tcaacgttgc tgacaaaggct 360
gcacgcacca aggaagtggc caagatcggtt gtcacatca ttggcgttca cactgtttt 420
gatcaacagg ctgttgcac gacaccgtt gccgatctca accttgggtt cagccgttac 480
cttgggtgtt acattttccgtt agctgggtgg tgcgtggcaaa acaagggtttt 540
gatgcagggtt ccacaaattgt tgggtgttgc ggcgtatctt atgggtgttgc cgatctgtt 600
gctgctgtt ctgaaatcgtt cgcgtggcc aagggttacac aaaggcgttgg tggcgtttt 660
ggctggctgtt agaaaactgtt cagttttt 687

```

SEQ ID NO: 155 moltype = DNA length = 546
FEATURE Location/Qualifiers
source 1..546
mol_type = genomic DNA
organism = Methylobacillus flagellatus

SEQUENCE: 155

```

atgaataat atcaagagct cgtggtaac aagctgacca atgtaatcaa taacacagca 60
gaaggcgtatc atgacaagat cctgacatg gtggatgtt cgggtcgat cttccgttgg 120
gttgctgttgc gttccctgtt ggtttccggc tttttccca tgcgtctcgat ccatgcaaggc 180
tatcaagtca gcatgggggg tgagggtgttca actccggatca tccagggtgg cgacctttt 240
atttgtatctt cccgggtgggg aaggccggaa accctgtatcc ccttgggttgcgaaaggccaaag 300
agccaggccggc caaaagggtcat cgtgtatccat atgaaaggccgc agtccccatggcggagctg 360
gcacacccatgg tgggtggat cgggtggcaatc gatgcgtatccatccacaa gacacatggc 420
atgcctatggt gcaatattttt cggatgttgc acgtctgttgc tccgtggaaac tacgttgc 480
aagtgtatcatcaaaaagg cctgacacaa gaaaggcatgc gtcgttgcattca cgccaaatctt 540
gaataaa 546

```

SEQ ID NO: 156 moltype = AA length = 693
FEATURE Location/Qualifiers
source 1..693
mol_type = protein
organism = Methylobacillus flagellatus

SEQUENCE: 156

```

MLPVKLTPY FINRELGILA FNRRVLAQAE DERVPLLERL KFLCIFSSNM DEFFEVRVAG 60
LNQIYKVNSP HVGADGMLPK QALAKISKLA HEMVDHQYKI LNEIVLPQLA AQEIRFIRRG 120
DWNEVQREWI ATYFFRELMP ILTPIGLDPA HPFPVRVLNKS LNFAVELEGK DAFGRNSGVA 180
IVQAPRALPR VIRLPRDIAG CDHGFVFLSS ILHAHVNLFS SGMTVKGCYQ FRVTRDSELT 240
LDDEETKDLR ALQGELTHR HYTPSLPRLW ADSCPAEMSA FLLQGQFGLGQ DDLYQVNGLV 300
NLVRLMOPID SVDRPDLKYP HYTPSLPRLW ADSCPAEMSA FLLQGQFGLGQ DDLYQVNGLV 360
IKQAAEDPQV LAIKQTFTYRT SADSTLMTSL IDAAQRGKEV TVVSELLARF DEEANINWAA 420
KLENAGAHVV YGVVGHKTHA KMAMVRRDE DKLRRYYHIA TGNYHORTAR LYTDGLLTC 480
QEEISEDVND VFAQLTGLGK ASKLRLHWWQS PFTLHQRTVK AIQNEAEIAR SGKKARIIAK 540
MNALLDPDV RALYDASCAG VQIDLIVRPGV CALRPGIPCGI SENIRVRSSIV GRFLEHTRIF 600
YFYDDGAEHII YLSSADWMYR NFFRRIEVCF PLLDAKVKRR VFKEGLEPYL KDNSNAWEMQ 660
SDGHYERKAS RRASFAAQOF LMAEHGQEIL PES 693

```

SEQ ID NO: 157 moltype = DNA length = 2082
FEATURE Location/Qualifiers
source 1..2082
mol_type = genomic DNA
organism = Methylobacillus flagellatus

SEQUENCE: 157

```

atgttgcggc tcaaaattgttgc tcccgatgtt ttcatcaatc gcgagcttggg catttcgtcc 60
ttcaaccgttc cgcgttggc ccaaggcttgc gacgacggcg taccctgtt tgcgttgc 120

```

-continued

aagttccctc	gcattttcag	cagaatata	gatgagtct	tggagggtcg	cgtacgcaggc	180
ctcaagaacc	agatcaaata	caactcgccg	catgtaggcg	ccgacggcat	gtgtccccaa	240
caggcgctcg	ccaagatcg	caagttggcg	catgaaatgg	tggatcacca	gtacaagatc	300
ctcaatgaaa	tctgtactccc	acagttggca	gccaggaaa	tccgcttat	ccgcggcggt	360
gactggaacg	agggtccagcg	cgagttggatt	gtctacttact	tcttcggcga	gttgatgccc	420
atttctaccc	cgatcggtct	ggatccagca	caccgttc	ctcgctactc	gaacaaaagc	480
ctgaactttg	ccgtcgagct	ggaaggcaag	gacgccttc	ggcgcataatc	gggcgtcgcc	540
atcggtcagg	ccgcggcgcc	gctgcgcgc	gtcatccgc	tgccggcgca	cattggcgcc	600
tgcgaccatg	gtttcgatt	cctgtcatcc	atctgtcag	cgatgttaa	cgaactgttc	660
tccggatcg	cggtaaaaagg	ttgttacccg	tccggcgatc	cgcgcgatag	cgagttgacc	720
ttggatcgat	aagagaccaa	ggacttcgtc	ctggccctgc	aagggtgagct	gacttcacgt	780
caatacggc	atgcgtacg	ccttggaaat	gcccacgt	gcccggcgga	aatgtccgccc	840
tttttactgg	gtcaatttgg	cctggggccag	gacgacatct	accagggtca	tggtctggtc	900
aacctctgtt	ggctgtatcg	gatccccggc	agcgtggacc	gcccggact	gaagtatccg	960
cattacacgc	ccaggctggc	acggggatgc	gataaggac	aggacatctt	caaggccatc	1020
cgcaggatcg	atatctgtct	gcatcatca	tcccaatcg	tcaatccggt	agtagaaatc	1080
atcaaggcagg	ctgcccgg	tccgtcaatg	ttggccatca	agcagacgtt	ctaccgcacc	1140
agcgcgcagact	ccacatcgat	gacatccctc	attgcgcggc	cacgcggccg	caaggaaatc	1200
accgtgggtgg	ttggatcggtt	ggcacgttt	gtatggaaat	ccaatataaa	ctggggcagcc	1260
aagctggaaa	atgcggcgcc	tcacgtcg	ttggcgatgg	tggggcaca	gaccatggcc	1320
aagatggcca	ttgttggcg	gcgcgtatgg	gacaatgtgc	ggcggttatgt	gcatattggc	1380
accggcaact	accaccaggc	tacccgcgc	ctgtatacc	acttcggct	gtcacccatgc	1440
caggaaagaaa	tcacggcg	tgtgtatgc	gttgcgtccc	agtcgtgggg	cctggggcaag	1500
gtctagcggc	tcgcgcattct	tcggcgtcg	ccgttccatct	tgccacccgc	cacggtc	1560
gcaatccaga	atgaagcgga	aattggcagg	teggggcaaga	aggccccat	cattggcaag	1620
atgaatgcgc	tactcgatcc	cgacgtgtc	cgacgttgc	acgtatgc	ctggccggcc	1680
gtgcacatcg	acccatcgat	gcgcggcg	tgcgccttc	gcccggat	tccgggcata	1740
tcggagaata	tcggggatcg	ttccatcg	ggacgatttc	tcgagcacac	gcccgcatttc	1800
tatTTTATG	acgacggcg	agaacacatt	tacccgtcg	gcccggact	gatgtatcg	1860
aatttctcc	ggccatcg	agtttgc	ccgtactcg	atgccaatgt	aaagaagcgc	1920
gtgttcaagg	aaggcgtgg	accatatctc	aaggataa	gcaacgcctg	ggaaaatgcag	1980
tccgatggcc	attacgaaac	caaggccag	cgccggcc	gttttgc	acagcaattc	2040
ctgatggcg	agcatggaca	ggaaaatttgc	cccgagatgc	ga		2082

SEQ ID NO: 158 moltype = AA length = 393
 FEATURE Location/Qualifiers
 source 1..393
 mol_type = protein
 organism = *Methylobacillus glycogenes*

SEQUENCE: 158
 NLVRLMQIPD SVDRPDLKYP HYTASVPKEL GKDDADIFKAI RKQDILLHHP FQSFNPVIEF 60
 IKQAAEAPQV LAIKQTFYRT SADSTLMKSL IEAAQRGKEV TVVVELLARF DEEANINWAA 120
 QLENAGAHVV YGVVGHKTHA KMAMVVRDE DKLRRYVHIA TGNYHORTAR LYTDGLLTC 180
 QEEISEDVND VFAQLTGLGR ASKLRLWQS PFTLHQRLIK AIQHEAEIAK SGKKARIIAK 240
 MNALLDPDVI RALYDASNAG VQIELIVRGV CALRPGIPGI SENIKVRSIV GRFLEHTRIF 300
 YFYDNGAEHI YLSSADWMYR NFFRRIEVCF PLLDAVKKR VFKEGLEPYL KDNINAWEML 360
 SDGSYESKPS RRAAYSAQQA LMTELQGEVL PEI 393

SEQ ID NO: 159 moltype = DNA length = 1182
 FEATURE Location/Qualifiers
 source 1..1182
 mol_type = genomic DNA
 organism = *Methylobacillus glycogenes*

SEQUENCE: 159
 aatctggatc gctgtatcgca gatccccatg agcgtggacc gcccggatctt caagtacccg 60
 cactacaccg ccaggcgatcc caaggaaatcg ggcaaggatgc cccatattttt caaggccata 120
 cgcaggatcg atatctgtct gcacatcccc tttcgtatcat ttaatccggt gatgttattt 180
 atcaaggcagg cggcgaaaga tccgcggatgc ctggcgatca agcagatctt ctaccgcacc 240
 agtgtgactt ccacatcgat gaaatccctt attgtatggcc cgcaggccgg caaggaaatc 300
 accgtgggtgg ttgatcgatc ggcgcgtttt gatgtatcgaa ccaatataaa ttggggccgc 360
 caatcgatcg atgcggcgcc ccatgtggatc tacggatgg tggggcaca gaccatggcc 420
 aagatggcca ttgtatcgatc ccgcgtatgg gataatgtgc gacgtatgtt gatgtatcg 480
 actggcaattt accaccaggc taccgcgcgc ctctacccg actttggctt attgtatcg 540
 caggaaagaaa tcacggcgaa cgtcaatcgat gtgtttgcac agctgtatgg cctgggtcg 600
 gccaggcgcc tcgtgcaccc ctggatcgatc ccctttatctt tgcatcaaccc ttgtatcg 660
 gcccattatcg atgaggccgc gatgtccaaatc tctggcaaga aggcgcgtat tatecgccaa 720
 atgaatgcgc tgctggatcc ggtatgtata cgtgcgtat acgtatgcctt caatggccgg 780
 gtacagatag agttgtatcgatc ggcgcgggtt tgcgccttgc gcccgtatgg tccggccatt 840
 tccgagaaata tcaagggtcg ctcaatgttgc tggggcgttcc ttgtatcgatc acgtatcg 900
 tattttctacg acaatggatc cgacgtatcc tacccgtcgatc gtgcggatgg gatgtatcg 960
 aacttctcc ggcgcatttc gatgttgc tccgtgtcg atgccaatgtt gaaaaagcgc 1020
 gtgttcaagg aaggcgtggc gccttaccc aaggacaata tcaatgtatc gggaaatgtt 1080
 tcggatggca gctatgatc caagcccgc cgtcgatcgcc ctactatgc gcaacaggcg 1140
 ctgatgacag aacttgggc gaaatgttgc cccgagatgc 1182

SEQ ID NO: 160 moltype = AA length = 693

-continued

FEATURE	Location/Qualifiers
source	1..693
	mol_type = protein
	organism = <i>Methylobacillus rhizosphaerae</i>
SEQUENCE: 160	
MLPVKLNPNEY FINRELGILA FNRRVLAQAE DERVPLLERL KFLCIFSSNM DEFFFEVRVAG 60	
LKNQIKYNSP HIGPDGLLPK QALATVSSMA HEMVELQYKI LNDVVLQPOLA AEEIRFIRR 120	
DWNNDTQREWI AAYFFRELMP ILTPIGLDP AHPFRVNLNS LNFAVELEGK DAFGRNNSGV 180	
IVQAPRALPR VIRLPREIAG CDHGFVFLLS ILHAHVSELF SGMTVKCYQ FRVTRDLDI 240	
LDDEDTTDLR LALQGELAHR QYGDAVRLEV ADNCPAEMSS FLLGQFGLTQ EDLYQVNGLV 300	
NIVRLMQIPD SVDRPDLPK LAHTEDIFKAI RKQDILLHHP FQSFNPVIEF 360	
IKAQAAEDPQV LAIKQTIFYRT SADSTLMKSL IEAAQRGKEV TVVSELLARF DEEANINWAA 420	
KLENAGAHVV YGVVGHKTHA KMAMVVRREE DKLRRRYVHIA TGNYHQRTAR LYTDGVLTC 480	
QEERISEDVND VFAQLTGLR ASKLRLHWQS PFTLHQRMVK AIQNEADIAR SGKKARIIAK 540	
MNALLDPDV1 RALYDASTAG VQIDLIVRGV CALRPGIPGV SENIRVRSIV GRFLEHTRIF 600	
YFYDNGAEHHI HLSSADWMYR NFFRRIEICF PLLDAKARR VFKEGLEPYL KDNTNAWEMQ 660	
QDGSYALKPN RRASFTAQQF LMQEHGQEI PDV 693	
SEQ ID NO: 161	moltype = DNA length = 2082
FEATURE	Location/Qualifiers
source	1..2082
	mol_type = genomic DNA
	organism = <i>Methylobacillus rhizosphaerae</i>
SEQUENCE: 161	
atgttgcggc ttaaatgaa tccgaatac tttatcaatc gcgaaactcgg cattctagcc 60	
ttcaatcgcc gcgtgtcggc gcaaggcagag gatgaacggg ttcccattt ggaacgcctg 120	
aagtttctcg qcaatttcg cagcaatcgat gatggattt tccgggtcg ctgtggccgt 180	
ctcaaaaacc agatcaagta caactctcca cacataggcc cagatggcct cctgccccaa 240	
caggcaactgg caaacatcgag cagcatggcg cagagatgg tccggcttca gtacaaaat 300	
ctcaacatcgat tgggtctgcc gcaactggcg gcagaagaaa tccgcttat ccggccgtggg 360	
gactggaaacg atactcgacg cgaatggattt tttccggcga gctgtatggct 420	
atccctacggc ctatcgcc gggccggcatacatttc cgcgcgtact caacaagagc 480	
ctcaatttcg cgggtqaact ggaaggcaga gatgtttcg ggcgttaattc cggcgtagcc 540	
atttgtcagg cacccgcgc cttaccgcgc gtcatccgcg tgccacgcg gattgcccgc 600	
tgtgaccatg gtttgtatt cctgtctcc atctgtcatc cccatgtcg cgaattttc 660	
tccggcatga cggtgaaaggg ttgtccaaat tccgtgtca cgcgcgcacac tgacctgatt 720	
ctggatgatg aagacaccac cgacccctagg ttggcttgc agggtqaact ggctcatcg 780	
cagtacggcg acgcggtacg cctggaaatgta gcaagataact gcccggcaga gatgtctcc 840	
ttcctgtctg gocaattttg gttcacacag gaagacctgt atcaagtaaa cggctgtgtc 900	
aaccttggta cctgtatcgca gattccggac agcgttgcg gccctgactt caagtttgc 960	
cactacacac cgatgtcgcc gaaaactg ctcaggaaatg cggatatactt caaggcaatc 1020	
cgcaaacagg atattctgtc gcatcatccg ttccagtcat tcaatctgt gattgaattc 1080	
atcaaaccagg cggcaaaaga ccctcaggatc ctggcaatca agcagacgtt ctatcgatc 1140	
agcgcagact caacgcgtat gaaatcattt attgaagcagc cgcgcgcgg ccaagaagtc 1200	
acagtgggtt tggactgtct ggcgcgttc gatgaagaaat ccaatattaa ttggccgcgc 1260	
aagcttgaga atccgggtgc ccatgtatgc tacgggtcg tggccacaa gacacacgc 1320	
aagatggcoca ttgttgcgtcc cggcggaaatg gataagctgc gtcgtatgtt acatatcgcc 1380	
acaggcaattt accaccatcg gacagcacgg ctgtatccatg atttggggg actgacatcg 1440	
caggaggaga tcagtgagga cgtcaatgtatc gatgtccgcg agctgacccg acttggccgc 1500	
gccagcaaac tgcgcacatc ctggcaatcc cggccatct tgcatacaatg catggcaatc 1560	
gccattcaga atgaagctg cattggccgc tccggcaaaa aagccgtat tateggcaatc 1620	
atgaatcgcc tgcgtatgc tgacgtatc cgtgcgtatc acatgtatc cggccgcgtt 1680	
gtgcaatcg acctatcgatc gctgtggcgtc tggtcatttc gccccggcat tcccgccgtc 1740	
tcggaaaata tcagggtgcg ctccatcgatc ggcgttttc tccggccatc cccatggcc 1800	
tattttctatg acaatggggc agaaacacatc catctttccca gtcgcacactg gatgtacccgc 1860	
aacttttccg cccgtatcgatc gatgtgttc cccgtactgg atgcggaaatg gaagcgacgt 1920	
gtattcaagg aagggtcgatc accttaccc tccggccatc ccaacgcgtt ggatgtcg 1980	
caggatggca gttatcgatc caagccaaac cggccggccatc gtttactgc acagcgttgc 2040	
ctgtatcgatc aacatggatca gggaaatattt cccggacgtt ga 2082	
SEQ ID NO: 162	moltype = AA length = 729
FEATURE	Location/Qualifiers
source	1..729
	mol_type = protein
	organism = <i>Methylobacterium organophilum</i>
SEQUENCE: 162	
APAEPVGAA RESVLEAGRS LRHTPERFVN RELSWLQFN RVLEESNPN HPPLERLRF 60	
SISANNLDEF PMVRVAGLID QVRAGITVPS QDGLSPSEQL VRIGAEVSLR AADQQARWRE 120	
LREELHAADI DIVEPAALTA EQAAWLEEFY LNHVFPVLTP LAIDPAHPPF FIPNLGTTIA 180	
LMLVRPKDGR LLRALIRLPG VIDRFVRLPV TESETAARLM PIEQVIGMFT GRLFPGYLVK 240	
GSGAFRVVRD SDLETEEEAAT DLVLFETAL KQRRRGVVR LEVEAGMSED LRSFVADELE 300	
IAAADAIFLVE GMLALNELSQ VVGIDRPDLK FKPYNPRFPE RIRESGGDFV AAIRQKDFIV 360	
HHPYYESFDV VQFLAQAAARD PNVVAIKQTL YRTSSNSPIV AALAEAELG KSVTALVELK 420	
ARFDEEANIR WARNLEKAGA QVVFVGFVELK THAKLSLVVR REGDKLVTYC HVGTGNYHPI 480	
TARIYTDSLFTADPAIARD VSRIIFNFIITG YAEPAELEMV AVSPLTLKNR LLEHIEAEVA 540	
HAKAGRPAAI WIKCNSLVDG QIIDALYDAS GAGVQIDCVV RGICCLRPGI PGLSETIRVK 600	

-continued

SIVGRFLEHG RIYAFGNGVG LPHPKATVYI SSADLMSRNL DDRVEALLPI TNPTVHQQL	660
DQIMLANLLD NRQSWRLLAS GGCEPNAHKY FMTNPSLSGR GKASKKSSPR	720
ALSRRAQRS	729

SEQ ID NO: 163 moltype = DNA length = 2190
 FEATURE Location/Qualifiers
 source 1..2190
 mol_type = genomic DNA
 organism = Methylobacterium organophilum

SEQUENCE: 163

```

tcaatgtcgat tggcggcgctc ggctgaggcc cccggggctc gacttttcg aggcccttgcc 60
ggctcccgag aggctggat ttggcatgaa atacttgtgg gcatttaacg gctctcgcc 120
ctcggcgccgc ttggatccgtt cgcagccgcg cgaggcgagg agccggcagc tctgacgggt 180
gtcgagaaagg ttggcccgatc tgatgtggc gggcacccgc tgatgtggc tcgggttcgt 240
gatcgccgcg acgcgcctcgac cacggcggtc gggatccggc ctcatcagggt cggccggagga 300
gatgttagacc gtggccctgg ggtggcccgatc cccgacgcggc ttggccaaagg cgtatgtcg 360
cccgtgtcg aggaaagcgcc cgacgatcgatc cttcacccggg atcgatctccg acaggccggg 420
gatggcccgcc cgccaggccgcg agatgccccgg gaccacgcgg tcgatgtggaa cgccggctcc 480
cgaggcgccgt tagagccgtcg cgatgtcgatc gggatccggc aaggagttgc acttgcattca 540
gatcgccggcc gggccggccgg ctttcgcgtg gggccatccgc gctctcgatc gctccaggag 600
ggcggttcc acgcgtcaggcg gcgagaccgc catgcgtcc agttcgccgcg gcttcggcgta 660
ggcggtatcg aagtggaaatc tgccgcacatc gtcccgccgcg atcgccggat ccggccgtgaa 720
gaacgcacagg tccgtgtata tgccgcgggt gatccgggttgc tagttgcggg tgccgcgtg 780
gcagtaggttgc agaggatgttgc cggccatccgcg tcggccacccgc aggcgcacgt tggcgtcg 840
cttgcgtcg acgaagccgcg acacgacccgt cgccggccgc ttctcgatc tgccgcggcca 900
gcggatgttgc gtttccctcgatc cgaacgcgcgatc cttcagatcgatc acggaggccgg tcacggact 960
gcggatgtcg cggccatccgcg cgaggccgcg gacgatcgatc gaggatcgacg aggtcgccgt 1020
gagcgtctgc ttgatcgatc cgcacgttcgg gtcgcgcgcg gcctggccgc ggaactgcac 1080
caccgcgtcg aaggattcgatc aggggtgggtg gacgatcgatc tccttcgtgc ggatcgccgc 1140
gaagacgtcg cggccgtctcgatc cgcggatcgatc tcggggaaagg cggccgggttgc agggcttgaa 1200
cttcaggatcgatc cggccatccgcg tgccgcacgcg ctggggacacgg ctttccggggccgc 1260
cttcacccgggg aagatcgatc cgcggccgtatc tcggccgcgcg agctgcgcgc 1320
atctccgcgcgatc cgcacgttcgg gacgatcgatc acggccgcgcg ggcgtcgatc 1380
caacgcgtcgc tcgaaaatgcgcgatc ggacgaggatc ttccgcgtcc tcctcgatc ccaggatcgatc 1440
gtcgcggacc acgcgcgtccgcg cggccgtccgcg ttccacgcggatc tagccggggaa agacgcggcc 1500
ggtaacatcgcgatc acatcgatcgcgatc ctcgcgtatc gtcgcgtatc ggcgcgttgc 1560
caccggcaggcc cgcacgcgcgatc ggtcgatcgatc gccggccgcgcg cggatcgacgcgc 1620
ccttcgcgtcc tcggccgcgcgatc caagatcgatc cgcgtatcgatc tcggggatgaa 1680
cgggaaacggggatc tgccggatcgatc cgcggccgcgcg cggccgtccgcg accggaaaggatc 1740
gaagatcttc cccatcgatc cgcgtatcgatc tcggccgcgcg cggccgtccgcg 1800
gatgtcgccgcgatc cgcgtatcgatc tcggccgcgcg cggccgtccgcg 1860
gaggccgcgcgatc cgcgtatcgatc cgcgtatcgatc tcggccgcgcg 1920
cgacggcaccgcgatc tgccggatcgatc cgcgtatcgatc cgcgtatcgatc tcggccgcgcg 1980
gaattcgatcgatc cggatcgatc tcggccgcgcg cggccgtccgcg 2040
gttcggatcgatc cgcgtatcgatc tcggccgcgcg cggccgtccgcg 2100
gttcacacgcgatc cgcgtatcgatc tcggccgcgcg cggccgtccgcg 2160
ggccgcgcgcgatc cgcgtatcgatc tcggccgcgcg 2190
  
```

SEQ ID NO: 164 moltype = AA length = 788
 FEATURE Location/Qualifiers
 source 1..788
 mol_type = protein
 organism = Methylococcus extorquens

SEQUENCE: 164

```

SSEVESDVVA RDTEATETEE TSLATERGV SESARARTAR AARILAAEKA AEKPSEADAK 60
VESDAAEAAR EPVLEAGRSL RHPSPRFVNR ELSWLQFNRR VLEESANPNH PLLEQLRFLS 120
ISANNLDEFF MVRVAGLHDQ VRAGLVPVPSQ DGLTPSEOLRA RIGTEVSRLA LDQQRWRAL 180
REELNGNGIH LAEPAEASAE HLAWLDDYFL TYIFPVLTPL AIDPAHPPFP IPNLGFTVAL 240
MMPVRPRDRGT LRALIRLPSL LDRFVRLPDV AGDGTARFIA IEQVIAKYAS RLFPGYSLKG 300
QGAFRVVRDSD LEIEEEAED LVLHFDSEAIIK RRRRGVVIRL EVEAGMTEEL RSFVADELEI 360
APDSVFVVEG MLALNELSQI VGDIPRDLKF KPYNPRFPER IRESGGDVFIA AIRQKDFIVH 420
HPYESFDSSV QFLAQAAARDP NVVAIKQTLY RTSSNSPIVA ALAEEAEAGK SVTALVELKA 480
RFDEEANIRW ARNLEKAGAQ VVFGFVELKT HAKLSMVVR EGDRRLVYCH VGTGNYHPIT 540
ARIYTDLSFF TADPAIARDV SRIFNFIITYG AEPAEALERMA VSPLTLKPRLL QHIEEEIAH 600
AKAGRPAIAW KICNSLWDQS IIDALYDASQ EGVQIDCIVR GICCLRPGIP GLSDTIRVKS 660
IVGRFLEHER VVAFGNGAGL PHPKATLYIS SADLMQRNLDRRVESLLPIT NPTVHQQLD 720
QIMLANLLDN RQSWRVLASG ASERISPADG EEPFNAHKYF MTNPNSLSGRG KSSKKSSPRA 780
LSRRAQRS 788
  
```

SEQ ID NO: 165 moltype = DNA length = 2367
 FEATURE Location/Qualifiers
 source 1..2367
 mol_type = genomic DNA
 organism = Methylococcus extorquens

SEQUENCE: 165

```

ttgtcgaaatcgatc tggatcgatc cggatcgatc cggatcgatc cggatcgatc 60
  
```

-continued

```

accaggcctcg ccaccgagcg ggggtgtgcgc agcgaatccg cccgcgcggc gacggcccgc 120
ggccgcgcga tcctcgccgc ggagaaggcc gcccggaaagc cgtcgaggc ggtgcgcagg 180
gtggagagcg atgcggcga cggcgcgcgc gacggcggtgc tggagggtgg gcgatcgctc 240
cgccactcgcc cgagcggtt cgtgaaccgt gagotgtctt ggctccaggta caaccgcgg 300
gttctggagg aatcgccaa ccccaaccat ccgtgtcgaa agcagetcg cttectctcg 360
atctcgccoca acaatctggc cgagtcttc atgtgcgtgc tgccggcct gcacgatcag 420
gtcccgcccg gcctcgctg gcccgcgcg gacggcctga cccctccga cgacgttgcc 480
cgatcgccca cggaaatctc cgggtggcg ctggaccaggc aggatcgctg ggcgcgcgtg 540
cgcgaggagc taaacggcaa cggttattcac ctgcggcgcg cggggaggt gtggcgagg 600
caccttgcgtt ggctcgacga ctatccatc acctacatcttcc tcccggtgt gacggcgtg 660
gcgtcgatcgcg gggccaccc gttccgttcaatcccaatc tgggttcaatc cgtcgcgctg 720
atgtgttgc gggcccggtt cggggcaccgc ctgcgcgcctc tgatccgcgtt gccgtcgctg 780
ctcgaccggc tcgtgcgcct gcccgaatgtt gccggcgacgc gcaatgccttgc 840
atcgagcagg tgatcgccaa atacggcgcg cggcttcttc cccgtactc actgaagggg 900
caggggcgtt tcgcgtcgatc gegegatccatc gatctcgatc tggaggaggaa ggccggaaagac 960
ctcgcttcgcg attcgaatc ggcgtatcggc cggccgcgcgc gggcgctgtt categcctg 1020
gagggtcgagg cggggcatgac ggaggagctg cgctccctcg ttgcgcacga attggagatc 1080
gccccccgatt cgggttcgtt ggtcgagggg atgtgtggcgc tcaacgactgt gtcgcagatc 1140
gtcggttgcgtt accggccggaa cctcaatgtt aaccggcttaca accccggctt ccccgagggc 1200
atccgttataat cggggcgccgat cgttccgcg cggatccgcg agaaggattt categtccac 1260
catccctacg atgcgttgcgat ctcgggtgtt cagttctcgat cccaggccgc cccgcgttcc 1320
aacgtgttgcgatc gacgtatcgcgat cgcacccatcgat cgaactccgcgatcgtggcc 1380
ggtcgatcgagg aggccggcgaa gggccggcgatc tccgttcccg cactgtgtca gtcgaaggcc 1440
cggttgcgttcc accggccggaa tateccgttgcgatc gtcgcgcgcgatcggcc 1500
gtgtgttgcgtt gtttcgttgcgatc gtcgaagactt caccgttgcgatc tggatgtgtt ggtggcgcc 1560
gaggggcgacc gtcgtcgatc ttcgttgcgatc gtcggccaccgc gcaactatca cccgtatcacc 1620
gcgcgcatctt acaccgcgtt ctcgttgcgatc accggccgcaccgc cggccatcgccgc 1680
tcgcgcgttcc tcaacttcatc caccggctac gcccggccgcg cggagatcgatc ggcgcgttcc 1740
gtgtcgccgcg tggccgttgcgatc gggccggccgcg ttcgttgcgatc tggaggaggaa gatcgccatc 1800
gccaaggccgcg gggccggccgcg gggccgttgcgatc atcaatgttgcgatc ttcgttgcgatc gtcgcgcgc 1860
atcatcgatcgttcc ttcgttgcgatc ttcgttgcgatc gggccgttgcgatc ttcgttgcgatc 1920
ggcatcgatcgttcc gggccgttgcgatc ttcgttgcgatc acacgtatcgatc ggtcaatgttgcgatc 1980
atcgatcgccgcg gtttcgttgcgatc gtcgcgcgttcc ttcgttgcgatc ttcgttgcgatc 2040
ccgcacccgcg aggccaccctt ctatattcc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2100
cgacgcgttccg attcgttgcgatc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2160
cagatcgatcgttcc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2220
ggcgagcgccgat ggtatcgccgcg ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2280
atgacgaatc aagccgttgcgatc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2340
ctcagccgcgttcc gggcccgatcgttcc ttcgttgcgatc ttcgttgcgatc ttcgttgcgatc 2367

```

SEQ ID NO: 166 moltype = AA length = 688
 FEATURE Location/Qualifiers
 source 1..688
 mol_type = protein
 organism = Escherichia coli

SEQUENCE: 166
 MQQEKLVIEK ELSWLSFNER VLQEAADKSN PLIERMRFLG IYSNNLDEFY KVRFAELKRR 60
 IIISEEQGSN SHSRHLLGKI QSRVLIKADQE FDGLYNELL EMARNQIFLI NERQLSVNQQ 120
 NWLRHYFKQY LRQHITPILL NPDTDLVQFL KDDYTYLAVE IIRGDTIRYA LLEIPSDKVP 180
 RFVNLPPEAP RRRKPMILLD NILRYCLDDI FKGFDFDYDAL NAYSMMTRD AEYDLVHEME 240
 ASLMELMSSS LKQRLLTAEPV RFVYQORDMPN ALVEVLRREKL TISRYDSIVP GGRYHNFKDF 300
 INFPNVKGKAN HINFKVPLRRL HIWFDAIRERD NGFDAAIRERD VLLYYPYHTF EHVELLRQA 360
 SFDPSPVLAIK INIYRVAKDS RIIDSMIHAH HNGKKVTVVV ELQARFDDEA NIHWAKRLTE 420
 AGVHVIFSAP GLKIHAKLFL ISRKENGEVV RYAHIGTGNF NEKTARLYTD YSLLTADARI 480
 TNENVRRVFNF IENPYRVPVT DYLMVSPQNS RRLLYEMVDR EIANAQOGLP SGITLKLNNL 540
 VDKGLVDRRLY AASSSGVPVN LLVRGMCMSL PNLEGISDNI RAISIVDRYL EHDRVYIFEN 600
 GGDKKVYLSS ADWMTRNIDY RIEVATPLLD PRLKQRVLDI IDILFSDTVKA RYIDKELSN 660
 RVPRGNRNRK VRAQLAIYDY IKSLEQPE 688

SEQ ID NO: 167 moltype = DNA length = 2067
 FEATURE Location/Qualifiers
 source 1..2067
 mol_type = genomic DNA
 organism = Escherichia coli

SEQUENCE: 167
 atgggtcagg aaaagctata catcgaaaaa gagctcaggat ggttattcgat caatgaacgc 60
 gtgcgttcagg aaggccggcga caaatctaaatcc cccgttgcgtt aaaggatgcg tttctcgcccc 120
 atctattccatcataaaccatcata tggatcttgcgtt aaaggatccgc tccgttgcgtt caatcaacaa 180
 atcattatttgcgaaaca aggctccaaatcc ttcgttgcgtt ccgttgcgtt gggcaaaattt 240
 cagtcggccggg ttcgttgcgtt ccgttgcgtt ttcgttgcgtt ttcgttgcgtt gtcgttgcgtt 300
 gagatggccgcg gcaaccatcata ttcgttgcgtt aatgaacgcg agtctccgtt caatcaacaa 360
 aactggctgcg ttcgttgcgtt taagcttgcgtt ctgcgttgcgtt acattacgcg gattttacatc 420
 aatccctgaca ctgcgttgcgtt gtcgttgcgtt aaaggatgtt acacctatcc ggcgttgcgtt 480
 attatccgttgcgtt gtcgttgcgtt ccgttgcgtt ttcgttgcgtt ttcgttgcgtt ttcgttgcgtt 540
 cgctttgttgcgtt attaccgcgcg gcaaccatcata ttcgttgcgtt acgttgcgtt ttcgttgcgtt 600
 aacattctgcgtt ccgttgcgtt ttcgttgcgtt ttcgttgcgtt ttcgttgcgtt ttcgttgcgtt 660

-continued

aatgcctatt	caatgaagat	gaccgcgtat	gccgaatacgtt	attttagtgca	tgagatggaa	720
gcccggctgtat	tggagttgtat	gtcttccagtat	ctcaaggcgcgtt	tttaacttcgtt	tgagccgggtt	780
cgtttttgtttt	atcagcgca	tatggccaaat	gccccatggttt	aatgtgttacg	cgaaaaaaactgtt	840
actatattcc	gotacgactc	catcgcccc	ggccgtcgat	atcataattt	taaagacttt	900
attaatttcc	ccaatgtcggtt	caaagccaaat	ctggtaaca	aaccactgcc	gcgttacgc	960
catatttttgtt	ttgataaaagc	ccagttccgtt	aatggttttt	atggcatttcgtt	cgaacgcgtt	1020
gtgttgtctt	attatccctt	teacacccgtt	gagcatgtgtt	tggaaactgtt	cgctcagggtt	1080
tcgttcgacc	cgagcgtactt	ggcgataaaat	attaacattt	accgcgtggc	gaaagatctt	1140
cgcacatcatcg	actcgatgtat	ccacgcgcgtt	cataacggta	agaaagtac	cgtgtgtgtt	1200
gagttacagg	ccgcgtttcgat	cgaaagaaagcc	aacatttcactt	gggcgaaaggc	cctgaccgaa	1260
gcaggcgttgc	cgctttatctt	ctctggccgtt	gggtgtaaaat	ttcacgcgtt	actgttctgtt	1320
atttcacgtt	aaaaaaacgg	tgaatgggtt	cggttatgcac	acatccccgtt	cggttacgtt	1380
aacgaaaaaa	ccgcgtgtt	ttatactgtat	tattcgtgtt	tgaccgcgtt	tgccgcgtat	1440
accaaacaaag	tacggcggtt	atttaaactt	atggaaaacc	cataccgtcc	ggtgcatttt	1500
gattattttaa	tgtgtatgtcc	gcaaaactcc	cgccgcctat	tgtgtatgtt	ggtgacccgtt	1560
gagatcgccaa	ccgcgcgtt	agggtgtcc	agtggatctt	ccctgttacgtt	aaataacttt	1620
gtcgatataag	gcctgggtt	tcgctgttat	gccccctccat	gtccggcggtt	accggtaat	1680
ctgtgtgttc	cgccgtatgtt	ttcgctgtat	cccaatctgg	aaggcatttt	cgacaacattt	1740
cgtgcgtatca	tgtgtgtt	ccgttacattt	gaacatgtac	gggttttat	ttttggaaaat	1800
ggccggcgata	aaaagggtcta	cctttttccgtt	gcccgttgcgtt	tgacgcgtttt	tattgttattt	1860
cgtatttgaag	ttggcgacgtt	gctgtgtat	ccgcgttgcgtt	agcagegggt	actggacatc	1920
atcgacatatt	ttttcagcgat	tacggtcaaaat	gacacgttata	tcgatataaga	actcgttataat	1980
cgttacgttcc	cccgccggca	tcgcgttgcgtt	gtacggggcg	atgtggcgat	tttgcgttttt	2040
atcaaatcact	tgcgttacacc	tgcgttacacc	atgcgttacacc	tttgcgttttt	tgcgttacacc	2067

SEQ ID NO: 168	moltype = AA	length = 249				
FEATURE	Location/Qualifiers					
source	1..249					
	mol_type = protein					
	organism = Methyllobacillus flagellatus					
SEQUENCE: 168						
MIDRQSSLGT	RPGVHQLYVS	LARNASEIAE	AQRLRYKVFA	EEMGANVSGT	EGLDQDGFD	60
FCDHLLVRNH	ETGEVVGTYR	ILSPAMANEA	GGYYSAGEFD	LSRLRHLAPS	MVEVGRACVH	120
KDYLRLGGTT	LLWAGLANYM	KVHRYEYIMIG	CGSIGMADGG	HMAASLYHQL	KQDHLSPEVY	180
RVFPNCPLPL	EALRNDELA	CPPLIKGYLR	LGAYICGAPA	WDPDFNTADM	LVMLPLSRLN	240
KRYAAHFFK						249

SEQ ID NO: 169	moltype = DNA	length = 750				
FEATURE	Location/Qualifiers					
source	1..750					
	mol_type = genomic DNA					
	organism = Methyllobacillus flagellatus					
SEQUENCE: 169						
atgtttgtacc	ggcagtcgttcc	cctccggacc	aggccccgggtt	tccaccatgtt	ctatgtcgtt	60
cttgcacggca	acgcacatccgtt	aatcgccgtt	gcacaaacgtt	tgcgttacaaat	agtatttgca	120
gaggagatgg	cgcccaacgtt	ctccggactt	gaaggactgg	accaagatgg	ctttgacgtt	180
ttttgcgtat	acttgcgttgc	ccgcgttacattt	gaaacagggtt	aatgtgttgcgtt	cacttacccgtt	240
atattgtat	cgccgttacgtt	ccacaaatgtt	ggccgttactt	acttgcgttgcgtt	cgttacgtt	300
cttcgtggcc	tgcgttacattt	tgcgttacattt	atgtgttgcgtt	ttggccggcggtt	ctgcgttacat	360
aaggatttttt	gcctgggggtt	cacgttacattt	ctgttctgttt	ccgggttgcgtt	caactacat	420
aaagtgcgtt	gttacgttgcgtt	catgttgcgtt	tgtgttgcgtt	tgcgttgcgtt	agacggccgtt	480
catatggcggtt	ctatgttgcgtt	tcaccatgtt	tcacgttgcgtt	tgcgttgcgtt	atgttttgcgtt	540
cccggttattttt	ccatgttgcgtt	gttacgttgcgtt	gaaacgttgcgtt	gcaacgttgcgtt	ggaagtgcgtt	600
tgcccgcccccc	tgatcaaaat	tttacgttgcgtt	ctgggttgcgtt	atatctgttt	cgccaccatgtt	660
tgggacccgtt	atttcaataat	agcgttacgtt	ctgggttgcgtt	taccattgtt	gctgttacgtt	720
aaacgttacat	cgttacattt	tttcaataat				750

SEQ ID NO: 170	moltype = AA	length = 248				
FEATURE	Location/Qualifiers					
source	1..248					
	mol_type = protein					
	organism = Methyllobacillus flagellatus					
SEQUENCE: 170						
MILERLSPKA	AEDDAPLHIS	FARNPSEVAE	AQRLRYRIFV	EEMGANLPSK	DGYDRDGFD	60
FCDHLLVRNN	NGEVVGTYRI	ILSPFMANEA	GGYSAEFDL	SRLSPLFDRT	VEVGRACVHE	120
DYRHGGTT	LLWAGLAKYMQ	ANRYYEYIMIGC	GSVGMGDGH	MAASLYNRLR	EDYLSPAEYR	180
VFPRVPLPLD	ALRSDMPATI	PPLMKGYLRL	GAYICGEPAW	WDPDFNTADM	LVMLPLSRLN	240
RYAAHFFK						248

SEQ ID NO: 171	moltype = DNA	length = 747				
FEATURE	Location/Qualifiers					
source	1..747					
	mol_type = genomic DNA					
	organism = Methyllobacillus flagellatus					
SEQUENCE: 171						
atgcttggaaat	ggcggtttat	acccaaggca	gcggaggacgtt	atgctccgtt	gcacatgttt	60

-continued

tttgcggcga	accgcgtctga	agtcgctgaa	gcgcagcgcc	tgcgcgtaccc	tatctttgtt	120
gaagaaaatgg	gtgccaacct	gcccggcaag	gacggctacg	accgtgacgg	ctttgacgct	180
ttctgcgacc	acttgcgtgt	acgcaacaad	aacggcgaag	tggcgccac	ctatcgata	240
tttagccccat	tcatggccaa	tgaaggccgt	ggctattact	cgccggccga	gttcgacttg	300
agccgttta	gcccactgtt	cgaccgtact	gtagaagtctg	gcgcgtccctg	cgtgcatgag	360
gattatcgcc	atggcgccac	catcaccatg	tttgtggctg	gcttggccaa	atacatgca	420
gccaacccgtt	atgaatacat	gataggttg	ggcagcgtag	gcatgggtga	cggtggacac	480
atggcagcoca	gtctttataa	tagatgcgt	gaagactatt	tgtcgccccg	tgaataccgc	540
gtgtttccac	gtgttccgt	gcccactggat	gcttggcga	gcgcacatgcc	agccaccatc	600
ccgccccatgta	tgaaaaggcta	cctggcactg	ggtgttccata	tctgtggcga	accggccctgg	660
gatcctgact	tcAACACCCG	tgacatgtct	gtgtatgttg	cgctgtccgc	tttggacaaa	720
cgttacgcag	cgcacttttt	caataaa				747

SEQ ID NO: 172 moltype = AA length = 249
 FEATURE Location/Qualifiers
 source 1..249
 mol_type = protein
 organism = *Methylobacillus glycogenes*

SEQUENCE: 172
 MLDRQLPLGN KPGASQLYVS LARNASEIAE AQRLRYKVFA EEMGANLAGA NGLDQDGFDA 60
 FCDHLLVRNQ DTGEVVGTYR ILSPAMANEA GGYSSAAEFD LTRLRHLAPS MVEVGRACVH 120
 KDYRMGTTT LLWAGLANYM KVHRYEYIMIG CGSIGMADGG HMAASLYHQL QKEHLSPAEY 180
 RVFPNCPLPL EALRHQEVET CPPLIKGYLR LGAYICGAPA WDPDFNTADI LVMLPLARLN 240
 KRYAAAHFFK 249

SEQ ID NO: 173 moltype = DNA length = 750
 FEATURE Location/Qualifiers
 source 1..750
 mol_type = genomic DNA
 organism = *Methylobacillus glycogenes*

SEQUENCE: 173
 atgcttgcgttacc ggcagttacc cctaggcaac aaacccggag cgagccact ttacgtcagc 60
 ctggcccgca accgcctctga aattgcagaa gcgcagccgc tgcgttacaa agtatttgca 120
 gaagaaaatgg gtgccaatct ggctggtgcg aatggccctgg atcaagatgg ctttgcgtca 180
 ttctgcgacc acttgcgtgtt ggcgaatcaa gacacgggggg aagtggtagg cacttacccgc 240
 atattaagcc cccgcattggc caatagacca ggccgttattt attccgcgc tgaatttgac 300
 ctcacgcgtt tacgtcacct ggacccagc atgttagaa tggtggccgtc ctgcgtgcac 360
 aaggattacc gcatggccgg caccattact ttgtgtggg caggtctcgc caactacatg 420
 aaagtacata catgatggcaatc catgatggat tgccgcacca tagggatggc cgatggccgt 480
 catatggcgcg acaggctgtt tccacatggc cagaagagc acctgtgcgc tgctgaatac 540
 cgtgtgttcc ccaatttgcgc cttgcgcgtg gaacccctgc gtcacgacca agaagtgact 600
 tgccgcgcatt tgatcaaagg ctatggcgc ctgggtgcct atatttgcgg cgcacccgccc 660
 tgggacccctg atttcaatac ggccgatatac ctgggtcatgc taccattagc ccgactcaac 720
 aaacgcgtatg cagctcactt tttcaataatg 750

SEQ ID NO: 174 moltype = AA length = 249
 FEATURE Location/Qualifiers
 source 1..249
 mol_type = protein
 organism = *Methylobacillus rhizosphaerae*

SEQUENCE: 174
 MLDRQSSLGI KAGASQLHVS LARNPSEIAE AQRLRYKIFA EEMGANLSSS DGLDQDGFD 60
 FCDHLLVRNH ETGEVVGTYR ILSPAMANEA GGYSSSGEFD LSRLRHLASS MVEVGRACVH 120
 KDYRLGGTTT LLWAGLANYM KVHRYEYIMIG CGSIGMADGG HMAASLYQQL QNDHLSPEY 180
 RVFPNCPLPL EALRHDEVN CPPLIKGYLR LGAYICGAPA WDPDFNTADI LVMLPLSRLN 240
 KRYAAAHFFK 249

SEQ ID NO: 175 moltype = DNA length = 750
 FEATURE Location/Qualifiers
 source 1..750
 mol_type = genomic DNA
 organism = *Methylobacillus rhizosphaerae*

SEQUENCE: 175
 atgctcgtatc ggcagttttc ccttggcatac aaggccgggg ccagccattt gcatgtcagt 60
 ctggcccgca acccttcga aattgcagaa gcccacatc tacgttacaa aatatttgc 120
 gaagaaaatgg gagccaatct ttccagctcc gatggccctgg atcaggatgg ttttgcgtca 180
 ttctgcgacc acttgcgtgtt tcgcaatcat gaaacccggc aagtgtcgg cacctacccgc 240
 atccctcgtatc cccgcattggc caatggatc ggtggcttattt actccatccgg tgaattcgac 300
 ttgtcgccac tgcgcacccat tgcctcaagg atgtgtggaa taggtcggtgc ctgcgtgcac 360
 aaggattatc gactggccgg caccattacc ttgtgtggg caggccgtgc caattacatg 420
 aaggatcatc gatgtggatc catgatggc tgcggcgttgc ttggcatggc cgatggccgc 480
 cacatggccgg ccagectgtt tcagcaatca caaaacgacc acctatcccc accggaaat 540
 cgtgtatttc ccaatttgcgc cttggcactg gaagcattgc gccatgaccc cgagggtgaat 600
 tgccgcgcgc tgatcaaagg ctacttgcgt ctgggtgcct atatttgcgg cgctccgtcc 660
 tgggacccctg atttcaacac cgcagatcc ctgggtcatgc taccattgtc gagactcaac 720
 aaacgcgtatg cagctcattt tttcaagttaa 750

-continued

```

SEQ ID NO: 176      moltype = AA  length = 332
FEATURE
source          Location/Qualifiers
1..332
mol_type = protein
organism = Methylobacterium organophilum

SEQUENCE: 176
MVANLARGAY SALGAGWHKG VQWGVGVAPIA KLRRRAVPLP GVAGAGGGIDD VLSAERAIRL 60
PEPFEGRSLG RIGSLEVRLA TRKSEIRRAQ RLRYKVFYEE MSAVPSGLAA LSRRDVGDYD 120
AVCDHLLVLD HAAPPRKKPF VEPRPKVVGT YRLLRGEVSE RHTGFYSESE YDLAPLLAAQ 180
GHRLRLLELGR SCVLKPYRSK RTVELLWQCI YAVVLHHRID ALICASLEG TDPDRLLALPL 240
AFLHHHHARAP EGWRARALPE RAVAMDRMPR EAVDAKAALQ ALPPLIKGYL RLGATFGDGA 300
VIDRQFQTTD VFVTLPVEVI GARYRGHFAP AG 332

SEQ ID NO: 177      moltype = DNA  length = 999
FEATURE
source          Location/Qualifiers
1..999
mol_type = genomic DNA
organism = Methylobacterium organophilum

SEQUENCE: 177
atggtcgcca atctcgccccg cggaggcctac agcgcgcctcg gagccggctg gcacaaggc 60
gttcaatggg ggtgtcgccgc gcccgtatcgaa aagctgcgcg gccggggcaact cccgtcgccg 120
ggcgtcgccgc ggcggggccg aatcgacgcgt gtgtcgccgcg cggagcgccg gatccgcctg 180
cccgagccgt tccaggccgcg cgcctcgccg cgcatecgccgat gcctcgagggt ccggcgccg 240
acgcggaaagt ccgagatccg gcccggcccg cggctgcgcgt acaaggctt ctacgaggag 300
atgtcgccgg tggccctccgg gtcgcgcgcg ctgtccggccg gcgacgtcga cggctacgac 360
ggcggtcgcc accacatgtcgt cggttcgtcgac caccggggccgaa gaagccgttc 420
gtcgagcgccgc gcccgaaggt cgccgcgcacc taccggctcg tcgcggccgaa ggtcgcccgag 480
ccgcacaccg gttctattc cgagacgcgcg taccgcgcgcg ccgcgtcgct cgccgcgcag 540
gggcacccggc ggctcttggaa actcgccgcgcg tccctcgctgc tcaaggcccta ccggagcaag 600
cgacccgtcgcc acgtcgatgtc ggcggggatcc taccgcctacat tcgtccatcgac 660
ggcgtgtcgcc gtcgcgcgcg cttggggatgt accgcattccgcg accggctggc cctgcgcgtc 720
gccttcgtgc accaccatgc ccgcgcgcgcg gagggtcgcc gggccggccg gtcgcggag 780
cgggcgggtgg cgatggaccg gatgcgcgcgcg gaggcggtgg atgccaaggc cgccctgcag 840
ggcgtcgccgc cgctgtatcaa gggttacccgcg cgcctcgccgcg acacccctgg ccgacggccgcg 900
gtgatcgacc gccagtcgtcg cacaacgcgcgat gtgtcgatcgat ccgtccggatc 960
ggcgccgcgc accggggccca ttccgcgcgcg gcggggtgt 999

SEQ ID NO: 178      moltype = AA  length = 332
FEATURE
source          Location/Qualifiers
1..332
mol_type = protein
organism = Methylorubrum extorquens

SEQUENCE: 178
MLPNFSPREA TAAWKAGWET KVQAPFARFR QTFEGGDLAL PGGAPILLKN RAKRPLLEPG 60
LDPSLGRIGA LEVRLATTKS EIRRAQLRY RVFYEEMSAV PTGMAALKRR DVDAYDAICD 120
HLLVIDHAAT DAKPFRKARP KVVGTYRLR QDRAEAHPGF YSSGEYDLAP LLERHPGKRF 180
LELGRSCVLK PYRTKRTVEL LWHGWAYVQ HHRIDAMLGC ASLEGTDPER LALPLSFLHH 240
HARAPEAWRA RALPERYVAM DRLAKEAVDP KAALMALPPL VKGYLRVGAT FGDGAVVDRQ 300
FGTTDVLVVL PVSGIAARYI GHFGASADRH AA 332

SEQ ID NO: 179      moltype = DNA  length = 999
FEATURE
source          Location/Qualifiers
1..999
mol_type = genomic DNA
organism = Methylorubrum extorquens

SEQUENCE: 179
atgtcgccga atttttccccg cggcgaggcc acggccgcgt ggaaggccgg ctgggagacc 60
aagggttcagg ccgccttcgc ccgttcgcgc cagacatccgcg aggccggccgat cctcgccctg 120
cccgccgcgtc gcccgcgtc tctcagaagaa cggggccaaatc gcccgtcgatc cgaaccgggg 180
ctcgacccca gctcgccgcgat cggccggccg ctcgaatgcgcg gcctcgccac cacaatgtcc 240
gagatccgcgcg gggccgcgcgcg cctcgccgcgc acggaggatgc gtcggcggtg 300
ccgacccggca tggccgcgcgc cagaacgcgcgc gatgtcgatgc cctacgcgcgc gatctcgac 360
caccttcgcgc tcatcgaccgc cggccgcgcgc gacgcgcgcgc cttccgcaca ggccggccgc 420
aagggtggcgcgc gacactaccgc gtcgtcgccgc caggaccggccgc gggaggccgc tttcggttc 480
taactccctcg gcgatgtacgc ctcgcgcgcgc ctgcgtggacgc gccatccagg caagccgttc 540
ctcgaaactcg ggcgcctcg cgtgtcgaaat cgcgtaccgc ccaagccgcac cgtcgaaactc 600
ctgtggcaccgc gcatctggccgc tctatgtcgatgc caccacccgcgc tgcacgcgcgc gtcggcgatc 660
gccagccgtgg agggcaccgcgc tccggaggatc ctggcgctgc ctttgacgtt tcttcacccac 720
catgcccccgcc ccccccgcgc gttggccggccgc cggcccccgcgc cggaggccgcata cgtcgatcg 780
gaccggctgg cgaaggaggccgc ggcgcgtatccgc aaggccggccgc tgcgtggccgc gtcggccgtg 840
gtgaaggccgtt atctccgggtt cggggccaccgc ttcggcgacgc ggcgggtggt ggaccggccag 900
ttcggccacca cccgacgttct cgtgtcgatgc cccgttcgcgc gcatcgccgcgc ggcgtatccgc 960
ggccatccgcgc gcgccgtgc ggtatcgccgc gccgcctgatc 999

SEQ ID NO: 180      moltype = AA  length = 352

```

-continued

FEATURE	Location/Qualifiers
source	1..352
	mol_type = protein
	organism = Methylorubrum extorquens
SEQUENCE: 180	
MLSPRKTHR AASRDGIVST MLPNFSPREA TAAWKAGWET KVQAPFARFR QTFEQGDAL 60	
PGGAPILLKN RAKRPLLEPG LDPSLGRIGA LEVRLASTKS EIRRAQRRLRY RVFYEEMSAV 120	
PTGMAALKR DVDAYDAICD HLLVIDHAAAT DAKPFRKARP KVVGTYRLLR QDRABAHFGF 180	
YSSGEYDLAP LLERHPGKRF LELGRSCVLK PYRTKRTVEL LWHGIWAYVQ HHRIDAMLGC 240	
ASLEGTDPER LALPLSFLHH HARAPEAWRA RALPERYVAM DRLAKEAVEP KAALMALPPL 300	
VKGYLRLVGAT FGDGAVVDRQ FGTTDVLVVL PVSGIAARYI GHFGAGADRHH AA 352	
SEQ ID NO: 181	moltype = DNA length = 1059
FEATURE	Location/Qualifiers
source	1..1059
	mol_type = genomic DNA
	organism = Methylorubrum extorquens
SEQUENCE: 181	
atgttttcgc cgccgcaagag gacgcatecg cggggctccc gcgacggat cgtgtcgatc 60	
atgctgcggc atttttcccc gcgcgaggcc acggccgcgt ggaaggccgg ctgggagacc 120	
aagggttcagg cgcccttcgc ccggttccgc cagaccttcg agggcggtga cctccgcctg 180	
cccgccgggtt cgcccgatccct tctgaagaad cggggcaacg gcccgtgtc cgaaccgggg 240	
ctcgacccca gccttcggcc gatcgggccg ctgcgaagtgc ggctcgccct cacaagtc 300	
gagatccgcc gggcgccageg ctgcgtcact cgcgtgttc acgaggagat gtcggcggtg 360	
ccgacccggca tggccgtct caagggccgc gatgtcgatg ectacgacgc gatctcgac 420	
caccttcctcg tcatecgacca cgccggccacc gacgcgaacg cttccgc当地 ggccggccgg 480	
aagggtggtcg gcacccatccg gtcgttgcgc caggacccggc eggaggggca ttccggcttc 540	
tacttcctcg ggagatcgca ctcgcgcgcg ctgcgtggccg gccatccggg caaacgc 600	
ctcgaactcg ggcgcctctg cgtgtcgaag ccttaccggca ccaagcgac cgtcgaactc 660	
ctgtggcacg gcatctgggc ctatgtgcag caccaccgcg tcgacgcgtat gctcgctgc 720	
gccaaggcctgg aggccacccgcg tccggagcgt ctggcgctgc cttccacccac 780	
catgcccccgcc ccccccggggc tgccggccgcg cggaggcgtca cgtcgcgatg 840	
gaccggctgg ccaaggaggc ggtcgaaaccg aaggccgcgc tgatggcgtt accggcgctg 900	
gtgaagggtt atctccgggt cggggcgacc ttccggcagc ggcgggtgtt ggaccgc 960	
ttccggcacca cccgacgtgtc ctgtgtgtc ccgtgttcac gcatcgccgc ggcgtacatc 1020	
ggccatttcg ggcgggtgc ggtatcgccac gccgcctgc 1059	
SEQ ID NO: 182	moltype = AA length = 288
FEATURE	Location/Qualifiers
source	1..288
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 182	
MTLAMNELAD ELCLIDVNET FLDNHVMDFE NAFPHKRMYR GSYPDLKDSS IVIITAGVPN 60	
RSDVKSRSNAY LQENISIFRK IGGVIETYVP NAIIVTASNP VDLLNYLYK AFDFOREQLI 120	
GYSLNDSYRF EWALRSTLGI EAKEHVVSPV IGEHGGSQVP LFHVRKSGE QLSVTPQQEE 180	
QIRAEKTWF VRFNELNVPR TTGWTGVM SKLVAKLSKD EPMETIGSAI VSGHYGLDDM 240	
SFGVPIKVNR NGIQQIEWD LSDQETKALT QSAKTIQAMI VENKAYFL 288	
SEQ ID NO: 183	moltype = AA length = 333
FEATURE	Location/Qualifiers
source	1..333
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 183	
MSKTKNKADN IIIIGDGAVG STFAYTLLRE NISVSIGIID PSQKKVAADV EDLCHTLPFT 60	
KSCHQSVVIA DYSDVKHEADL VVITANAPQA KFGEDKDRLG LLQQNVRMMT QITTDIMRNG 120	
FDGVILVVSN PVDALTQVQ RVSGLLTSRV IGTGTLLDTI RLHALIANKL RVNPNDIKGF 180	
IIGEHGNNSGF INWSHLTIVS IPIILWLEKY TSYDCYQQLF QELEEVVRKI GVDIFADKGN 240	
TSYGIASCLV MITETILSEN NRILPVSVYL TGEYGIDGVY IGAPVRLNRQ GISQIIIDL 300	
NESELAFAFKQ SAEILQANIL ALELTDELAL LVE 333	
SEQ ID NO: 184	moltype = AA length = 318
FEATURE	Location/Qualifiers
source	1..318
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 184	
MAIKHGNKVS VIGGAGAVGA TVSVVLTQSG LIQDVLVLDI AKEKTEGEAL DLNHGASFID 60	
PMNIRAGGYE DTKNSDIVII TAGIAQKPDQ TRIELVEKNI EVMKDIIPREV TKYSPNSIIL 120	
LISNPVDILS YIAYKLSGFP KGRIIGSGTV LDTSRKFPEI GKHFQVDPRD VETVYLGEHG 180	
DTSFPAWSLT NIKSIPISEY AELMGIEFNE EFKYQAYENV KQAAVDVINK KGATYYAIGL 240	
SAMHIVESIL KDMRRILPVS TLINDYYGAD DLYLGMPCIV GRNGVVEVLK INLSEEIQN 300	
VNKSANALTT VLNESFRK 318	
SEQ ID NO: 185	moltype = AA length = 317

-continued

FEATURE	Location/Qualifiers
source	1..317
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 185	
MPQHHQKVVL IGDGAVGSSY AFAMVQQGIA EELVIVDVAK EQTVDGALDL EDATAWTYNK 60	
NVRSGDYEDA KDAELAVITA GAPQKPGETR LDLVDKNLKI MQSIIKPLMA SGFDGIILVA 120	
ANPVDILITMA AQRFSLPLTN RVFGSGTSLD TARLRVNLLKG EFGVAPESVD AYILAEHGDS 180	
EFANFAEATI GGRPLMDWAA EKGLTEDDLE QILYDTAHLKA YEIIINRKGAT FYGIGTSLAR 240	
ITRAIFRDER AVLPGAYLD GQYGLNDIYI GTPAIIGANG VESVIETKLT DGELERFHAS 300	
AKTLKEVAEG GFAKLAD	317
SEQ ID NO: 186	moltype = AA length = 314
FEATURE	Location/Qualifiers
source	1..314
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 186	
MNTKARKVMI IGAGNVGTS AYALLNQTC EELILVLDLDR ARVQAHQDL SDAAAYMPGM 60	
MTISTREVD CADVIAVIT VSGGSLKQQC TRLDELNATA KIVNSIVPQM MDNGFNGLFL 120	
VATNPDCIIT WQVWKLSGLP RSQVIGTVW LDTTRLRRVL AQALDGAQS IDAFILGEHG 180	
DTCFPVWSHS SVYGSPIADV YLKHTGQRDL RQQIAEKVRT LGFEIYAGKG CTEYGVAGTI 240	
AEICRNFITG SHRALAISCI LDGEYGYMDV AVGVPAVLTO SGVQQIIELQ MADDEAAQFR 300	
RSVEVIKANI ARLG	314
SEQ ID NO: 187	moltype = AA length = 260
FEATURE	Location/Qualifiers
source	1..260
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 187	
MSAIHFSTYK IFHVNLLENL ASCDIIILAF RKESLKEPLS RLVELQNNIL ELKDIVLTLK 60	
NANFKGKYIV ATNPNDTITY YTQVLSQLPK NHVFGSGTML DSSRLKKLLA KDLNINSKDI 120	
FACMIGEHGD SQFAALSNAS VLGNQNLDFY KQLGKDLDI QELEKAVISE GFYIYERKGR 180	
TEFGIGTSC T NLAKAILKDR KSLHPVSVIF DDIAFSMPAI IGKDGIEKVF ELKPNEKEKT 240	
KLENSKQQIK NAIQSVKDKI	260
SEQ ID NO: 188	moltype = AA length = 319
FEATURE	Location/Qualifiers
source	1..319
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 188	
MQLNQRKAAI IGCDFVGSAT AFCLMQSGLF SEIVLQDVDK DKAEGEAMDI THGTPFAGRM 60	
KIYAGNYNDM MDAAVIIITA GANQKEGETR LDLVKKNKKI FEEIIPEISN RDYEGILLIV 120	
SNPVDILTYT AIKISGFEEN RVIGSGTVLD SARFRYLLKG HLDVDSRSVQ AFIIGEHGDS 180	
EIAAWSSVNV SGIPVNDFCE MRGHYDHHEA MKAIAADNVKE SAYKIIKRKK ATYYGIAMSV 240	
KRICEVILKD EKAILPVSTM MHGAHGIEDI VLSMPAIVGK NGIETQVPIE LNEEEEERKLK 300	
QSARILKQMI GIKKDDVVKK	319
SEQ ID NO: 189	moltype = AA length = 325
FEATURE	Location/Qualifiers
source	1..325
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 189	
MLQPTNAKNH QKIVLVGAGA VGSTFAYAT LOGIQQELVI DVARDKAHG DAKDIADALL 60	
FTYPKNIYAG DYSDAADADI VVITAGTAQK PGESRLLDVN RNLHIMQSIV TKIVESGFNG 120	
IPLVASNPVD ILTYATWKLS GFPALERVGTS GTSLDSARLC KVVGYLNFNVD PRTVSGYMLG 180	
EHGDTEFPW SHLSVGGLTM AEWIKNDPRF TEADLDIAD RVVNAAYDII RLKGATYYGV 240	
GAALARICRA LLDDENTILP VSVYLNQYQG VEDMYIGTPA VLNRHGVROV IEIPLNEKEQ 300	
GLMRASASQL DEVMRNTFAE IGIQR	325
SEQ ID NO: 190	moltype = AA length = 318
FEATURE	Location/Qualifiers
source	1..318
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 190	
MMKIGVIGAG SVGVGICNYL LTMGSVSELV LLDQNLERAE GEVFDFRHTA ALTFSKNTHI 60	
IPFKDYLDD AADIVVITAG AQIKQGQTRI DIAEINAIG VDIARQVERV APNAILIVVS 120	
NPCDLVSHFI VSNTTFKPSK VISSGCVIDT ARLMTIVANR VQLDPKNVFG YVLGEHGSHC 180	
FTPCKSLISIA GQPADYYCDT NHIKRIDADE LLESVKQAGY EIFKRKHNTT HGISASFVRI 240	
IQAQIMINEKS VLPVGTMLSG QYGLDNVLM S LPTVVGKQGA EKVLMHPPFSD EELSTLARIA 300	
ENVTAVVNEV AQVTGLKA	318

-continued

SEQ ID NO: 191	moltype = AA length = 308
FEATURE	Location/Qualifiers
source	1..308
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 191	
MPVSSRIAIV GAGNVGATTA YALMLRGLFA EIVLIDTSD RATAEATDIA DANAIARPARI	60
IWAGDYADAR DASILVITAG AATADGEART SVAGKSAAV RTCVDAAMAA GFDGILVVAS	120
NPADAMAQVA QATSGLPAAR VIGTGTLDS NRFRKRVADQ LAIAPGAVEG LVLGEHGDSE	180
VMAYSSVRIG GLDLDDAYLDG AAFDRASVAH EVVRAGYTIS HGKGHTSFVG ATAIVRICEA	240
IQRDEHIVLP VSALKLGQFG ISNLYLSLPC LVGAAGIMRI LTPDLTTEET AALHASADAL	300
RRTLVCK	308
SEQ ID NO: 192	moltype = AA length = 316
FEATURE	Location/Qualifiers
source	1..316
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 192	
MVNEKKVAIV CGGFVGSSSA FALMQSGLFS EMVLIDVDKN RAEGEALDIA HGMTFAEPMK	60
IYAGDYSVDA DAAMIVVTAG AAQPKGETRL DLVNKNVNIF KSIPIEKKS GFDGILLIVS	120
NPVDVLTYYA IKMSGLPEGH VIGSGTVLDT GRLQQMLGAH VEVDPRDVQA YVMGEHGDSE	180
FVAWSSAQVA GPVLNTFCEL HGHLEHETAЕ KRIAEDVKNS AYTIIIEKKHA YYGVAMAVK	240
RICTAVMRDE QTVLVPSSLM VGEYGLSDL AISMPTVVGRD GVVCRVPVPL NEGEKSELNA	300
SAEALKDIID SVDFSC	316
SEQ ID NO: 193	moltype = AA length = 323
FEATURE	Location/Qualifiers
source	1..323
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 193	
MARVEKPRKV ILVGDGAVGS TFAFMSMVQQG IAEELGIIDI AKEHVQGDAI DLADATPWY	60
PKNIYAADYP DCKDADLIVI TAGAPQKPGT TRLDLVNKNL KILSSIVEPI VESGFDGIFL	120
VVANPVDILT HATWKISGFP KDRVIGSGT LDTGRLQKVI GEMEHVDPRS VNAYMLGEHG	180
DTEFPAWSYN NVGGVKVSDW VKAHGMDESK LEDIHKEVAN MAYDIINKKG ATFYCIGTAS	240
AMIAKAILND EHRVLPLSVA MDGQYGLHDL HIGTPAVVGR KGLEQIIEMP LSDDEQAKME	300
ASAKQLKEVM DKAFKETGVK VRQ	323
SEQ ID NO: 194	moltype = AA length = 325
FEATURE	Location/Qualifiers
source	1..325
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 194	
MVTSRHSRRV VVGTGSGVSA VAGAIVLQGL CDELVLINHN PDKAEGLAMD LMDGSEFIGR	60
FVSIRRGDWS DCRDADVVIV TAGPKPKGE TRPQELGAII DIVDPILRAI TRSGFNGIL	120
MVSNPVDVMS WFAWRRTGLP RAQVLGSGTA LDTSRMKAII GEITRLDPRM VSGYVIGEHG	180
ESQFVPWSTV SFIGKSFTQY LQDNRERYEG VTTEVIEERT RERGMAIKAL RGRTSQGIAA	240
TVAGLTRTIL WNERRVIPVS TLIDGEYEYD EHDVFLSLPV ALDADGVGDF VDLHLTGDEL	300
AKPHESARVV REHCVLIADR LRCQS	325
SEQ ID NO: 195	moltype = AA length = 314
FEATURE	Location/Qualifiers
source	1..314
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 195	
MVMNNMRKVGI IGGVGHVGAHV ALSLAMQGIC DEIQLCDIDE EKIVSEQDLD LDAVCYLPHR	60
VRVTLGKYEE LADCDIIVSS VGKIDVQSVN RLSELQQSVD MIRTTFVPRLR KAGFQGIYLN	120
ITNPCDIIISN QIQKESGLPA ARVIGTGTGL DTSLRKAVLA RETGFDHKSI TTYMLGEHGD	180
SQMAAWSVTA FGGKPLAQLE EEEPERFGFD KEVLAEEVRK AGWKTLNGKG ATEFGIASTA	240
ARMISCIHFHD EKQIIPFISTRY LDGEYGEAGL YASVPVVLGK DGVDEIILKN LDDKEMGEFQ	300
KTC SVMKEHI ALIK	314
SEQ ID NO: 196	moltype = AA length = 321
FEATURE	Location/Qualifiers
source	1..321
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 196	
MKNNNHRKVVL IGDGAVGSSF AYALLHQGIV GELAIINHTL SKAEGDALDL EDAIPYLENI	60
KIYSAATYNT CKDADIVVIC AGAPQKPGES RLQLVTKNFK IITSITDKVI ASGFDFGIFIV	120
AANPVDIMSY AVLQSGFTP ERVIGTGTSL DSARLKIALS KKFEINPKNV DINILGEHGD	180
SEFAAAYSAAK IGGIPLLKLI SQNGLTLLTDL YNIEKSVRDK AYQIIDKKGA YYGVATAIM	240
KIVKAILRNE HEMITVGCYL NGQYGYNDVF VGTPAIMGNT GIERVIEWPL NSTEKNAMRK	300

-continued

TVGTLKKTTL NALNDLNLDT K	321
SEQ ID NO: 197	moltype = AA length = 315
FEATURE	Location/Qualifiers
source	1..315
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 197	
MKGNLNMSNK VLLVGDGRVG STFANDLLQN VDIDELVITD VAEKLPVGDS LDLEDVTNFY	60
KPVNIHAGTY ADAKDADIVV ITAGMARKPG MTRLDLVDKN VQILKSIIEP IVESGFKGIF	120
VVSANPVDIL TTLTQKLSGF PKNRVIGTGT SLDTARFKIA LAKKVGVTVR DVNAVVLGEH	180
GDTSFENFDE ATIDDKPLRS YPELTDEVLD DLQTEVRQKG CKIIANKGAT FYGVAQMLTQ	240
ICKAILENKE MVLPLSAPVK GLYgidHDLF LGTPAVINSN GIADVIETKL SDDELKKMNY	300
SADKMQEVD GVSLD	315
SEQ ID NO: 198	moltype = AA length = 322
FEATURE	Location/Qualifiers
source	1..322
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 198	
MALVSKIGII GTGHVGAHVA NALLFKGLAT ELYLSDLDEV LCKAQVNLL DAMPFYPHPA	60
RVFEVDTTRYE ELACDClIVN AAGHIEAAAAA SRDGELFVTT DEARTFAKRI SDAGFKGIWV	120
NIANPCDVVS TEIQYLTGCD PTRVIGSGTT LDSARFRHAL SVATGYPASC INAWNLGEHG	180
NGQFALWSHV TFGCLTDKEV EAQTGLKFDR AQLEQDARMG GYVTYKGKHC TEYSIANGAV	240
EVIGAIVNNT KLVTVPVSTLL DNVYVGASGFY SSLPAVIGKD GVEKVLVPEL SDNEIAAWKK	300
SCEHVKGNIQ QLDWLKVNDAR VE	322
SEQ ID NO: 199	moltype = AA length = 319
FEATURE	Location/Qualifiers
source	1..319
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 199	
MAQKHHHQKV VLGDGGMVGSA FGYALMQQGV AEELAIVDVA KDYATGDALD LEDAAPWTYP	60
KQVTGGHDYQ VVADADIVVI TAGVGRKPGQ TRLELIDKNL TIVKQVVDDNV MAQGFTGIFV	120
VATNPVDIIT LAVQQFSGLP EHHVIGTGTA LDTARLQVAL AEQYGVAPAT IDVVLVGEHG	180
DSAFANFDEA QIGGQSLNDF NKKYGNNSAQD LVELMEATTK KGGAIIGRKG ATFYRVATAL	240
ARIVRAILRD ESMVLPISAW MSGQYGLSDM YIGSPAVING DGAKTVIMAA LSPAEQMQMQ	300
RSAEILRAVT ADALANFMK	319
SEQ ID NO: 200	moltype = AA length = 308
FEATURE	Location/Qualifiers
source	1..308
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 200	
MSRKVFLVGD GAVGSTFAND LLQNAKVDEL AIFEVTKDPR VGDAMDLEDI TPFMGQTDIH	60
PADYPSDAKDA DVCVIAAGVP RKPGETRLDL VAKNVKILKS IVQPVVESGF KGVFVVSNSP	120
VDILTTLTK LSGFPKNRVI GTGTSLSMSR LRVELAKKLN VPVAKVNSMV LGEHGDTSE	180
NFDESTVDGK PLRDYAEIND DVLSEIETDV RKKGKIIAK KGATFYGVAM MLTQIVSAVL	240
DNRSICLPLS APINGEYGIK HDLYLGPTV INGEGIEQVI ETKLSDAEKA KMINSADKMQ	300
EVLGRIEL	308
SEQ ID NO: 201	moltype = AA length = 316
FEATURE	Location/Qualifiers
source	1..316
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 201	
MGIKSRKVAI VGTGLVGSSI GFSLITQGVC EEILMIDINE EKALGEAMDL NHCIYEYLDQH	60
TKVKVGKYED CGDVDAIIIT AGPPPCKPGKS RLDTLGASAM IMESIITPIM ESGFSGHFII	120
ISNPVDIMAY HWKQLSGLPK SHVIGTGTGV DSARLKHPIA DLFDVDPRSV QAYSMEHGD	180
SQMPWPWSHV IGGKPFNKVII NDNKHRVGD1 DLDKLVSDTA KAGWEVFNRK GTTYYGIATA	240
AVGVLKAIFN DENRIIPVST LLEGEYGFDG VFAGVPAVLN RNGVKELVEI EMTEEEKVKF	300
DKSVNLKDY CNTLKY	316
SEQ ID NO: 202	moltype = AA length = 317
FEATURE	Location/Qualifiers
source	1..317
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 202	
MNKFKGKVV LIGNGAVGSI YAFSLVNQSI VDELVIIDLD AEKVRGDVMD LKHATPYSPT	60
TVRVKAGEYS DCHDADLVVI CAGAAQKPGF TRLDLVSKNL KIFKSIIVGEV MASKFDGIFL	120
VATNPVDILA YATWKPSGLP KERVIGSGTI LDSARFRLLL SEAFDVAPRS VDAQIIGEHG	180

-continued

DTELPVWSHA NIAGQPLKTL LEQRPEGKAQ IEQIFVQTRD AAYDIIQAKG ATYYGVAMGL	240
ARITEAIFRN EDAVLTVSAL LEGEYDEEDV YIGVPAVINR NGIRNVVEIP LNDEEQSKFA	300
HSAKTLKDIM AAEELK	317
SEQ ID NO: 203	moltype = AA length = 309
FEATURE	Location/Qualifiers
source	1..309
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 203	
MNKVLLVGD GAVGSTFAND LLQHTSVDEL VICDVVKNRP VGDSMDLEDI LPFTGECNIH	60
SQEYSDAKDA AVVITAGVP RKPGESELRLD VNKNVSILKS IIQPVVESGF DGIFLVSANP	120
VDIRLTTLQK LSGFPKNKVI GTGTSLDTAR LKVDLARRLN VPVAGVNMSM LGEHGDTSFE	180
NFDESTVGGK PLTEFAHIDD NILNQVELNI RQKGKGIEN KGATFYGIAM MLTQIVTAIL	240
TNRSIALPLS APINGEYGIK SELYLGTPAV VDGSGISQVI ETELSESEKA KMVDSAAMQ	300
EVLNNIALD	309
SEQ ID NO: 204	moltype = AA length = 310
FEATURE	Location/Qualifiers
source	1..310
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 204	
MARTVGIIGS GHVGATLANN LLVTGSVDKL VLIDTNELKV NSDATDFEDA LANLPTHTKV	60
VKNNYRELKH ADVVVIAVGS IGVQNDDAKH DRFVELKVTS KAAKEVGTRL REVGFGVLV	120
SISNPNCDVVA ALFQHYTGLP RDQVIGTGTL LDTSRMKKV GORFELDPRS VSGYNLGEHG	180
NSQFTAWSQV RVLDQFVTAA LSEALDELA EAVERAGGYTA FHGKHYTNFG IAAAALRLVT	240
AIINDARAEV PVSNFREELG TYVGYPAVVG SGGVVRQPQL TLTSAGKEKL LASANYIQTR	300
YEAVLTELEK	310
SEQ ID NO: 205	moltype = AA length = 320
FEATURE	Location/Qualifiers
source	1..320
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 205	
MKSRSRKVAI VGAGMVGSSC AYSMVNQAIC DEIMMIDRTY DRAMAQALDL SHCMDFTSTR	60
TKVYAGTYSD CAKMDVILIT AGANPTAGQT RLDVLEASAV ITREIITNIM AGGFDFGIVVV	120
AANPVDIVTY MWKVISGLPR HRIIGTGTIS DSSRLKTLLS DVFSIDPRSV NGYALGEHGE	180
SQFVAWSHTV IGGKPILQIL DQHRERFPHL DLEDISRKTK DAGWEIFTRK GSTHFGIGSA	240
LAYITRSILN DEHKIIAVSA ILDGEYGHSG VCTGVPAIIG NTGIQELLEL NLNTEETQKL	300
NHSCNIVRAG IESLQLEDSI	320
SEQ ID NO: 206	moltype = AA length = 319
FEATURE	Location/Qualifiers
source	1..319
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 206	
MKTFGKKVVL IGDGSVGSSY AFAMVTOQVA DEFVIIDI AK DKVAKDVQDL NHGTVHSPSP	60
VDVKAGEYED CKDADLVVIT AGAPOKPGET RLQOLVEKNTK IMKSIVKSMV DSDFDGYFL	120
AANPVDIVLT FVKEYTGLPA ERVIGTGTIV DSARLQYLIS QELGVAPSSV DASIIGEHGD	180
TELAVWSSQAN VAGISVYDTL KEQTGSEAKA EEIYVNTRDA AYEIIQAKGS TYYGIALALM	240
RISKAILNNE NNVLNVSQQL DGQYGGHKGV YLGVPTLVNQ HGAVKIYEMP LSAEEQALFD	300
KSVKTLEDTF DSIKYLLED	319
SEQ ID NO: 207	moltype = AA length = 318
FEATURE	Location/Qualifiers
source	1..318
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 207	
MJQTRKVGIV GVGHVGSHCA LSMILQGVCD EMILMDIIP KAKAHAIIDCM DTISFLPHRA	60
IIRDGGIQEL SKMDIIVISV GSLTKNEQR EELKGSLEAI KSFVPDVVKA GFNGIFVTIT	120
NPVDIVTYFV RELSGFSKRN VIGTGTGLDS ARLKRILSEV TNIDSQVIQA YMLGEHGDTQ	180
VANFSSATIQ GVPFLGYMK HPEQFKGIEL SVLEKQVVRT AWDIISGKNC TEFFIGCTCS	240
NLVKAIFHNE RRVLPCSYL EGEYGHSGFY TGVPAAIGSN GIEEILELPL DERERKGFEA	300
ACAVMRKYIE IGKSYKIV	318
SEQ ID NO: 208	moltype = AA length = 316
FEATURE	Location/Qualifiers
source	1..316
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 208	
MNKLLRKVAI LGTGFVGSSST AYALINQGIC DELLIIDMKQ EKAVGESLDL IHCMDFLPSR	60

-continued

TKVYTGSIQN TADVDIIIVS AGPSPNENQQT RLDLTEGAAQ IMNDTIPRIM KTGFNGIFLI	120
ATMPVDTIITY HIWKLSGLAR HQVIGTGTSI DSSRLKTYLA QVLDIDVRSI QAYTMGEHGD	180
SQFAVWSHCT IGGKPPLLEIL DEKPHLATKL DLDDDTVEKVK RVGFEILKRK GTYYGGIGNA	240
LAYFTTQILN DSQQTLPASC ILDGEYGEKS IATGVPGKIS RNGIQDIVEV NLSEQEKQLF	300
RHSNTVLRDY MRRIGY	316
 SEQ ID NO: 209	
FEATURE	moltype = AA length = 318
source	Location/Qualifiers
	1..318
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 209	
MIILALERQK VALIGDGAVG SSYAYAMMQQ GLAEEFVIVD VVKERTVGDA LDLEDAQVFT	60
SPKHVYSGDY EDCHDADLAI ITAGAPQPKPG ETRLDLVNKN LKIMQAIIKP LVASGFKGVI	120
VVAANPVDIL YTAQQKFGSF PKNKVFGLSSGT SLDSARLVRV LSKKLNISQQ SVDAYLGEH	180
GDSEFAAYS AARVGGEPFLD VAKRTGLSDD DLATIEDQVR HKAYEIIINRK GATFYGVATC	240
LMRISRAILR DENAILPVGA PMNGEYGLND VYIGSPAVIN GSGIEKVIEV PLNDKEKAAM	300
KASAETLQKT TKDGMQDL	318
 SEQ ID NO: 210	
FEATURE	moltype = AA length = 320
source	Location/Qualifiers
	1..320
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 210	
MSSMPNHQKV VLVDGAVGS SYAFAMAQQG IAEEFVIVDV VKDRTKGDAL DLEDAQFTA	60
PKKIYSGEYS DCKDADLVVI TAGAPQPKPG ETRLDLVNKN NILSSIVKPV VDSGFDGIFL	120
VAAANPVDILT YATWKPSGFP KDRVIGSGTS LDSSRLRVAL GKQFNVDPRS VDAYIMGEHG	180
DSEFAAYSTA TIGTRPVRDV AKEQGVSDDED LAKLEDGVRN KAYDIINLKG ATFYGIGTAL	240
MRISKAILRD ENAVLPVGAY MDGQYGLNDI YIGTPAVIGG TGLKQIIESP LSADELKKMQ	300
DSAATLKKVL NDGLAELENK	320

1. A genetically engineered methylotrophic bacterium which has been modified to have an increased protein expression of a polypeptide having lactate dehydrogenase activity compared to an otherwise identical bacterium that does not carry said modification, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOS: 1 to 130 and 182 to 212, and ii) a polypeptide comprising an amino acid sequence, which has at least 70% sequence identity with the amino acid sequence of any one of SEQ ID NOS: 1 to 130 and 182 to 212, wherein said bacterium is of the genus *Methylobacillus*.

2-21. (canceled)

22. The bacterium according to claim 1, wherein the bacterium comprises at least one exogenous nucleic acid molecule comprising at least one nucleotide sequence encoding the polypeptide having lactate dehydrogenase activity.

23. The bacterium according to claim 1, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOS: 1 to 48; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70% sequence identity with the amino acid sequence of any one of SEQ ID NOS: 12 to 48.

24. The bacterium according to claim 1, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOS: 49 to 130; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70% sequence identity with the amino acid sequence of any one of SEQ ID NOS: 49 to 130.

25. The bacterium according to claim 1, wherein the polypeptide having lactate dehydrogenase activity is not allosterically regulated by fructose-1,6-bisphosphate.

26. The bacterium according to claim 1, wherein the polypeptide having lactate dehydrogenase activity has an amino acid not being histidine at the amino acid at position corresponding to *L. casei* LDH Histidine 188, as determined by sequence alignment.

27. The bacterium according to claim 25, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOS: 182 to 210; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70% sequence identity with the amino acid sequence of any one of SEQ ID NOS: 182 to 210.

28. The bacterium according to claim 25, wherein the polypeptide having lactate dehydrogenase activity is selected from the group consisting of: i) a polypeptide comprising an amino acid sequence of any one of SEQ ID NOS: 185, 186, 188, 189, 191, 194, 196 and 202; and ii) a polypeptide comprising an amino acid sequence, which has at least about 70% sequence identity with the amino acid sequence of any one of SEQ ID NOS: 185, 186, 188, 189, 191, 194, 196 or 202.

29. The bacterium according to claim 1, which has been further modified to have a decreased expression and/or activity of an endogenous polypeptide having polyphosphate kinase activity compared to an otherwise identical bacterium that does not carry said modification.

30. The bacterium according to claim 1, which has been further modified to have a decreased expression and/or activity of an endogenous polypeptide having Acyl-homo-

serine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

31. The bacterium according to claim **1**, which has been modified to have a decreased expression of an endogenous polypeptide having Acyl-homoserine-lactone (AHL) synthase activity compared to an otherwise identical bacterium that does not carry said modification.

32. The bacterium according to claim **1**, wherein said bacterium is *Methylobacillus flagellatus* or *Methylobacillus glycogenes*.

33. A method for producing lactate comprising cultivating a bacterium as defined in claim **1** under suitable culture conditions, which produce lactate.

34. A method for producing D-lactate comprising cultivating a bacterium as defined in claim **23** under suitable culture conditions, which produce D-lactate.

35. A method for producing L-lactate comprising cultivating a bacterium as defined in claim **24** under suitable culture conditions, which produce L-lactate.

36. A method for producing L-lactate, comprising cultivating a bacterium as defined in claim **25** under suitable culture conditions, which produce L-lactate.

37. The method according to claim **33**, comprising cultivating said bacterium under suitable culture conditions in a culture medium comprising a reduced one-carbon compound or a multi-carbon compound that contains no carbon-carbon bonds.

38. The method according to claim **37**, wherein the culture medium comprises methanol.

39. The method according to claim **33**, wherein the cultivation is performed in a bioreactor.

* * * * *