

pSTRAINS AND PROCESSES FOR SINGLE CELL PROTEIN OR BIOMASS PRODUCTION

FIELD OF THE INVENTION

5 The present invention relates to the production of protein and/or other macromolecules using microorganisms. In particular, the invention relates to novel bacterial strains and continuous culture processes for the production of protein or biomass using bacteria wherein gases and minerals are supplied to the cells. The invention also relates to the products of these processes and use of these products in e.g. food or feed.

10

BACKGROUND OF THE INVENTION

Growing world population, climate change and shortage of water increasingly pose a threat to traditional agriculture and thus sufficient supply of food and feed. Therefore, alternative sources of organic molecules, such as proteins, are being investigated. A potential alternative is single cell production, i.e. the production of protein and/or other macromolecules using microorganisms.

Chemoautotrophic microorganisms have been described which are able to grow on minimal mineral medium with hydrogen gas as the energy source and carbon dioxide as the only carbon source. For a review of these microorganisms, see e.g. Shively et al. (1998) Annu Rev Microbiol 52:191. Patent application WO2018144965 describes various microorganisms and bioprocesses for converting gaseous substrates into high-protein biomass.

25 However, various chemoautotrophic microorganisms have different properties in terms of growth rate, yield, biomass composition as well as properties related to being used as a food ingredient such as safety in human consumption, taste, smell, mouth-feel, technical and functional properties in cooking, etc. Not every chemoautotrophic microorganism has sufficient growth rate and provides sufficient yield and not every process can realistically be upscaled to an economically viable large-scale process. In order to have sufficient output of functional protein, e.g. for food or feed applications, it is important to find a suitable production organism and a suitable process which can be performed at large scale. This need is addressed by the present invention.

SUMMARY OF THE INVENTION

35 In a first main aspect, the invention relates to an isolated bacterial strain VTT-E-193585 or a genetically-modified or mutated strain generated using said bacterial strain,

wherein said genetically-modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.

5 In further aspects, the invention relates to a culture comprising the bacterial strain of the invention or a genetically-modified or mutated strain generated using said bacterial strain, wherein said genetically-modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source. Furthermore, the invention relates to a process for the production of biomass and/or protein, said process comprising culturing the bacterial strain of the invention or
10 a genetically-modified or mutated strain generated using said bacterial strain, wherein said genetically-modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.

BRIEF DESCRIPTION OF THE FIGURES

15 **Figure 1.** Optical density measured at 600 nm (black circles) and optical density probe readings during chemoautotrophic 200-L cultivation of isolated bacterial strain deposited as VTT-E-193585.

Figure 2. Optical density measured at 600 nm during parallel chemoautotrophic 200-mL cultivations of isolated bacterial strain deposited as VTT-E-193585 on different
20 nitrogen sources.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

When used herein, the term "isolated", e.g. in the context of a strain, means isolated
25 from its natural environment. Preferably, an isolated strain is pure, i.e. free of other strains.

The term "derivative", when used herein in the context of a strain, refers to a strain which is derived from a reference strain, i.e. generated using the reference strain as
30 starting point. E.g. a genetically-engineered or otherwise mutated or genetically-modified strain is an embodiment of such a derivative. Genetic modifications include point mutations, as well as insertions or deletions, including insertions or deletions of entire loci or fragments thereof. The derivative preferably has fewer than 10 genetic modifications, e.g. fewer than 5, such as 4, 3, 2 or 1 genetic modification(s) compared
35 to the reference strain.

When used herein, the noun "culture" refers to a suspension of viable cells in a liquid medium.

5 The term "biomass" has its usual meaning in the field of bacterial fermentation and refers to cellular material.

The term "continuous culture", when used herein, refers to a culturing process wherein fresh media is added continuously to the culture and media with bacterial culture is removed continuously at essentially the same rate.

10

Aspects and embodiments of the invention

In a first main aspect, the invention relates to an isolated bacterial strain VTT-E-193585 or a genetically-modified or mutated strain generated using said bacterial strain, wherein said genetically modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.

20 Strain VTT-E-193585 has been isolated from the seashore of the Baltic sea in Naantali, Finland. This organism is able to grow in suitable bioreactor conditions with minimal mineral medium with hydrogen as the energy source and carbon dioxide as the carbon source at limited oxygen conditions. 16S sequencing and Illumina metagenomics sequencing have shown that the strain most likely is a member of the genus *Xanthobacter*, but is not a known species. The bacterial strain is highly suitable for food and feed applications, because the dried cell powder has a high protein content and contains all the essential amino acids. It also contains more unsaturated than saturated fatty acids and a high level of B-group vitamins. The levels of peptidoglycans and lipopolysaccharides, which may cause allergy or toxicity, are low. A toxicity analysis was performed and no genotoxicity or cytotoxicity was observed for the strain. In addition, the strain is generally sensitive to antibiotics.

30

Strain VTT-E-193585 (SoF1) has been deposited on June 11th, 2019 in the VTT Culture Collection at the VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Finland, an International Depositary Authority under the Budapest Treaty. Further information on the characteristics of the strain and methods for culturing the strain are provided in the Examples herein.

35

In a preferred embodiment, if the strain is a derivative of strain VTT-E-193585, the derivative has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.

- 5 In one embodiment, if the strain is a derivative of strain VTT-E-193585, the derivative comprises the 16S ribosomal RNA set forth in SEQ ID NO:1 or a 16S ribosomal RNA having up to 20 nucleotide differences with SEQ ID NO:1, e.g. 1 to 10, such as 1 to 5, e.g. one, two or three nucleotide differences with SEQ ID NO:1.

10 SEQ ID NO:1. 16S ribosomal RNA sequence of strain VTT-E-193585:

```

CTTGAGAGTTTGATCCTGGCTCAGAGCGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGA
GCGCCCAGCAATGGGAGCGGCAGACGGGTGAGTAACGCGTGGGGATGTGCCCAATGGTACG
GAATAACCCAGGGAACTTGGACTAATACCGTATGAGCCCTTCGGGGGAAAGATTTATCGCCA
15 TTGGATCAACCCGCGTCTGATTAGCTAGTTGGTGGGGTAACGGCCACCAAGGCGACGATCA
GTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGTG
TGATGAAGGCCTTAGGGTTGTAAAGCACTTTCGCCGGTGAAGATAATGACGGTAACCGGAGA
AGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCG
20 GAATCACTGGGCGTAAAGCGCACGTAGGCGGATCGTTAAGTCAGGGGTGAAATCCTGGAGCT
CAACTCCAGAACTGCCCTTGATACTGGCGACCTTGAGTTCGAGAGAGGTTGGTGGAACTGCG
AGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCCAACTGGC
TCGATACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTC
CACGCCGTAAACGATGGATGCTAGCCGTTGGGCAGCTTGCTGTTCACTGGCGCAGCTAACGC
25 ATTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATTA AAACTCAAAGGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTGCAAGCAACGCGCAGAACCTTACCAGCCTTT
GACATGGCAGGACGATTTCCAGAGATGGATCTCTTCCAGCAATGGACCTGCACACAGGTGCT
GCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCC
TCGCCTCTAGTTGCCAGCATTAGTTGGGCACTCTAGAGGGACTGCCGGTGATAAGCCGAGA
30 GGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACA
ATGGTGGTGACAGTGGGATGCGAAAGGGCGACCTCTAGCAAATCTCCAAAAGCCATCTCAGT
TCGGATTGTA CTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCAT
GCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTT
TACCCGAAGGCGCTGCGCTAACCCGCAAGGGAGGCAGGCGACCACGGTAGGGTCAGCGACT
35 GGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCTGCGGCTGGATCACCTCCTTT

```

In a further aspect, the invention relates to a culture comprising the bacterial strain of the invention or a genetically-modified or mutated strain generated using said bacterial strain, wherein said genetically-modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source. In a preferred embodiment, the volume of the culture is 100 mL or more, e.g. 1 L or more, such as 10 L or more, e.g. 1,000 L or more, such as 10,000 L or more, e.g. 50,000 L or more, such as 100,000 L or more, e.g. 200,000 L or more.

In a further aspect, the invention relates to a process for the production of biomass and/or protein, said process comprising culturing the bacterial strain of the invention or a genetically-modified or mutated strain generated using said bacterial strain, wherein said genetically-modified or mutated strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source. In one embodiment, the process comprises culturing the strain in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide. Various further embodiments of the process are described herein below.

According to the genome sequence, the strain deposited under number VTT-E-193585 uses most likely Calvin-Benson-Bassham cycle for the carbon fixation where carbon dioxide molecule is connected to 5-carbon chain of ribulose 1,5-bisphosphate forming two molecules of glycerate 3-phosphate. This enables the strain to synthesise all the other organic molecules it requires for growth. Energy from hydrogen comes into the cell most likely through NAD^+ -reducing hydrogenases and/or NiFeSe-hydrogenases. In essence that is a redox reaction where hydrogen (H_2) is oxidized to H^+ and NAD^+ is reduced to NADH. In addition to ATP, NADH is one of the main energy carriers inside living organisms. Alternatively, some other energy equivalent is reduced by another hydrogenase enzyme using H_2 . The Calvin-Benson-Bassham cycle requires energy in the form of ATP and NADH/NADPH in order to fix CO_2 . The strain most likely generates ATP through oxidative phosphorylation, which consists of four protein complexes generating a proton gradient across a membrane. The proton gradient is generated using mainly energy from NADH. The proton gradient drives the ATP synthase complex generating ATP. According to the genome sequence, the strain has a bacterial F-type ATP synthase.

It is to be understood, when it is specified that the process comprises culturing the strain with an inorganic carbon source, that the inorganic carbon source is the *main*

6

carbon source in the culture. Thus, there may be minor amounts of organic carbon sources present in the culture, but the main metabolism and growth of the culture is based on the utilisation of the inorganic carbon source, preferably carbon dioxide, as carbon source. Preferably the proportion of the carbon supplied to the culture that is organic is less than 5%, such as less than 1%, e.g. less than 0.1% of all carbon supplied to the culture during the process. Preferably, no organic carbon sources are supplied to the process.

Similarly, it is to be understood, when it is specified that the process comprises culturing the strain with hydrogen (H_2) as energy source, that hydrogen is the *main* energy source in the culture. Thus, there may be other minor energy sources present in the culture such as ammonia, which may be supplied as nitrogen source, or minor amounts of organic compounds, but the main metabolism and growth of the culture is based on the utilisation of hydrogen as energy source. In the overall process hydrogen is preferably produced by water electrolysis; i.e. by splitting water with electricity to hydrogen and oxygen gases. Thus, the hydrogen and oxygen gases are provided to the bioreactor from an electrolyser nearby. Alternatively, electrodes may be placed inside the bioreactor to produce hydrogen and oxygen in the bioreactor rather than in a separate electrolyser.

The inorganic carbon source comprising carbon dioxide may comprise other inorganic carbon sources, such as e.g. carbon monoxide. In one embodiment, only carbon sources in gaseous form are provided to the culture. In a preferred embodiment, carbon dioxide is the only inorganic carbon source, and indeed the only carbon source, provided to the culture. In one embodiment, only gases and minerals are provided to the culture and the level of carbon dioxide in the gas provided is between 10% and 50%, e.g. between 15% and 45%, such as between 20% and 40%, e.g. between 25% and 35%, such as between 26% and 30%.

In another embodiment, gases and minerals are provided to the culture and the level of hydrogen (H_2) in the gas provided is between 30% and 80%, e.g. between 35% and 75%, such as between 40% and 70%, e.g. between 45% and 65%, such as between 50% and 60%.

In another embodiment, gases and minerals are provided to the culture and the level of oxygen (O_2) in the gas provided is between 10% and 25%, e.g. between 15% and 20%, such as between 16% and 18%. In another embodiment, the level of oxygen

provided is such that the level of dissolved oxygen in the culture is maintained at between 5% and 10%.

5 In a preferred embodiment, only gases and minerals are provided to the culture and the gas provided comprising H₂, CO₂ and O₂, wherein the percentage of H₂ is between 40% and 70%, the percentage of CO₂ is between 18% and 28% and the percentage of O₂ is between 12% and 22%.

10 Typically, the process of the invention includes the addition of a nitrogen source. The nitrogen source may for example be provided in the form of ammonium hydroxide, an ammonium salt, such as ammonium sulphate or ammonium chloride, ammonia, urea or nitrate, e.g. potassium nitrate. In other embodiments, nitrogen gas (N₂) is provided as nitrogen source. In a preferred embodiment, the nitrogen source is ammonium hydroxide or an ammonium salt, such as ammonium sulphate.

15 In one embodiment, the nitrogen source provided is ammonium hydroxide at a concentration of between 100 mg/L and 10 g/L, such as between 250 mg/L and 4 g/L, e.g. between 0.5 g/L and 2 g/L, such as between 0.75 g/L and 1.5 g/L.

20 Typically, the process of the invention includes the addition of minerals, such as minerals containing ammonium, phosphate, potassium, sodium, vanadium, iron, sulphate, magnesium, calcium, molybdenum, manganese, boron, zinc, cobalt, selenium, iodine, copper and/or nickel. Suitable mineral media are well-known art, and have e.g. been described in *Thermophilic Bacteria*, CRC Press, Boca Raton, FL, Jacob K. Kristjansson, ed., 1992, for example on page 87, Table 4.

In one embodiment, the minerals added include one or more of the following: ammonia, ammonium (e.g., ammonium chloride (NH₄Cl), ammonium sulphate ((NH₄)₂SO₄)), nitrate (e.g., potassium nitrate (KNO₃)), urea or an organic nitrogen source; phosphate (e.g., disodium phosphate (Na₂HPO₄), potassium phosphate (KH₂PO₄), phosphoric acid (H₃PO₄), potassium dithiophosphate (K₃PS₂O₂), potassium orthophosphate (K₃PO₄), disodium phosphate (Na₂HPO₄·2H₂O) dipotassium phosphate (K₂HPO₄) or monopotassium phosphate (KH₂PO₄); sulphate; yeast extract; chelated iron (chelated e.g. with EDTA or citric acid); potassium (e.g., potassium phosphate (KH₂PO₄), 30 potassium nitrate (KNO₃), potassium iodide (KI), potassium bromide (KBr)); and other inorganic salts, minerals, and trace nutrients (e.g., sodium chloride (NaCl), magnesium sulphate (MgSO₄·7H₂O) or magnesium chloride (MgCl₂), calcium chloride (CaCl₂),

calcium sulphate (CaSO_4) or calcium carbonate (CaCO_3), manganese sulphate ($\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$) or manganese chloride (MnCl_2), ferric chloride (FeCl_3), ferrous sulphate ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$) or ferrous chloride ($\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$), sodium bicarbonate (NaHCO_3) or sodium carbonate (Na_2CO_3), zinc sulphate (ZnSO_4) or zinc chloride (ZnCl_2), ammonium molybdate (NH_4MoO_4) or sodium molybdate ($\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$), cuprous sulphate (Cu_2SO_4) or copper chloride ($\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$), cobalt chloride ($\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$) or cobalt sulphate (CoSO_4), aluminium chloride ($\text{AlCl}_3 \cdot 6\text{H}_2\text{O}$), lithium chloride (LiCl), boric acid (H_3BO_3), nickel chloride ($\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$) or nickel sulphate (NiSO_4), tin chloride ($\text{SnCl}_2 \cdot \text{H}_2\text{O}$), barium chloride ($\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$), copper selenate ($\text{CuSeO}_4 \cdot 5\text{H}_2\text{O}$), sodium selenate (Na_2SeO_4) or sodium selenite (Na_2SeO_3), sodium metavanadate (NaVO_3), chromium salts).

In a preferred embodiment, the process of the invention includes the addition of one, more or all of: NH_4OH , KH_2PO_4 , $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, $\text{NaVO}_3 \cdot \text{H}_2\text{O}$, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, CaSO_4 , $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, H_3BO_3 , CoSO_4 , CuSO_4 , NiSO_4 .

In one embodiment, the medium provided to the cells comprises less than 1 g/L of chloride salts, such as less than 0.25 g/L of chloride salts, e.g. less than 0.1 g/L of chloride salts, such as less than 0.025 g/L of chloride salts, e.g. less than 0.01 g/L of chloride. In one embodiment, no chloride salts are supplied to the culture.

In another embodiment, no vitamins are supplied during the process, i.e. the media provided to the culture does not contain vitamins.

In another embodiment, no amino acids are supplied during the process, i.e. the media provided to the culture does not contain amino acids.

In another embodiment, no organic compounds are supplied during the process, i.e. the media provided to the culture does not contain any organic compounds.

In certain embodiments, the pH of the bacterial culture is controlled at a certain level. In certain embodiments, pH is controlled within an optimal range for bacterial maintenance and/or growth and/or production of organic compounds. In one embodiment, the pH in the culture is maintained between 5.5 and 8.0, e.g. between 6.5 and 7.0, such as at 6.8.

In certain embodiments, the temperature of the bacterial culture is controlled. In certain embodiments, temperature is controlled within an optimal range for bacterial maintenance and/or growth and/or production of organic compounds. In one embodiment, the culture is grown at a temperature between 25°C and 40°C, e.g. between 28°C and 32°C, such as at 30°C.

Typically, the process of the invention is carried out in a bioreactor. A bioreactor is utilized for the cultivation of cells, which may be maintained at particular phases in their growth curve. The use of bioreactors is advantageous in many ways for cultivating chemoautotrophic growth. Generally, the control of growth conditions, including control of dissolved carbon dioxide, oxygen, and other gases such as hydrogen, as well as other dissolved nutrients, trace elements, temperature and pH, is facilitated in a bioreactor. Nutrient media, as well as gases, can be added to the bioreactor as either a batch addition, or periodically, or in response to a detected depletion or programmed set point, or continuously while the period the culture is grown and/or maintained. In a continuous culture process, nutrient media, as well as gases, are added to the bioreactor continuously. Furthermore, bacteria-containing medium is being removed from the bioreactor continuously.

In a preferred embodiment, the volume of the bacterial culture is 100 mL or more, such as 1 L or more, e.g. 10 L or more, such as 100 L or more, e.g. 1,000 L or more, such as 10,000 L or more, e.g. 50,000 L or more, such as 100,000 L or more, e.g. 200,000 L or more.

In one embodiment, the productivity of the culture is more than 0.1 g cell dry weight per liter per hour, such as more than 0.2, e.g. more than 0.3, such as more than 0.4, e.g. more than 0.5, such as more than 0.6, e.g. more than 0.7, such as more than 0.8, e.g. more than 0.9, such as more than 1 g per liter per hour.

Bacteria can be inoculated directly from a cell bank, or via a seed culture at a smaller scale. Preferably, supply of fresh media to the culture and removal of used up media with bacteria is occurring at the same rate, such that the volume in the bioreactor remains the same.

In one embodiment, after an initial phase of reaching a suitable cell density, the bacteria grow at steady state or pseudo steady state, remaining continuously in their log phase,

at an OD600 above 5, such as above 10, e.g. above 20, such as between 50 and 200, e.g. between 50 and 100.

In one embodiment of the process of the invention, the bacterial strain has a growth
5 rate of 0.04 – 0.12 h⁻¹.

In another embodiment of the process of the invention, the liquid feed rate in the continuous phase is 50-80% of the growth rate.

10 In one embodiment, the bacterial strain used in the process of the invention is selected from the group consisting of: *X. agilis*, *X. aminoxidans*, *X. autotrophicus*, *X. flavus*, *X. tagetidis*, *X. viscosus*, *Xanthobacter sp. 126*, *Xanthobacter sp. 91* and strain VTT-E-193585.

15 In a preferred embodiment, the bacterial strain used in the process of the invention is VTT-E-193585 or *X. tagetidis*. Most preferably the strain used in the process of the invention is VTT-E-193585.

In another embodiment, the bacterial strain used in the process of the invention
20 comprises the 16S ribosomal RNA set forth in SEQ ID NO:1 or a 16S ribosomal RNA having up to 20 nucleotide differences with SEQ ID NO:1, e.g. 1 to 10, such as 1 to 5, e.g. one, two or three nucleotide differences with SEQ ID NO:1.

In another embodiment, the bacterial strain used in the process of the invention
25 comprises a gene encoding a ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) large chain having the sequence set forth in SEQ ID NO:3 or a sequence having more than more than 93% identity, e.g. more than 95% identity, such as more than 96% identity, e.g. more than 97% identity, such as more than 98% identity, e.g. more than 99% sequence identity to the sequence set forth in SEQ ID NO:3.

30

In another embodiment, the bacterial strain used in the process of the invention
comprises a gene encoding a ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) small chain having the sequence set forth in SEQ ID NO:5 or a sequence
35 having more than 83% sequence identity, e.g. more than 86%, identity such as more than 90% identity, e.g. more than 95% identity, such as more than 96% identity, e.g. more than 97% identity, such as more than 98% identity, e.g. more than 99% sequence identity to the sequence set forth in SEQ ID NO:5.

SEQ ID NO:2:

Nucleotide sequence of Ribulose biphosphate carboxylase large chain:

ATGGGTGCCGAAGCAACCGTCGGGCAGATCACGGACGCCAAGAAGAGATACGCCGCCGGCG
 5 TGCTGAAGTACGCCCAGATGGGCTACTGGAACGGCGACTACGTTCCCAAGGACACCGACCTC
 CTGGCGGTGTTCCGCATACCCCCCAGGCGGGCGTGGACCCGGTGGAAGCCGCCGCGGGCGG
 TCGCCGGCGAAAGCTCCACCGCTACCTGGACCGTGGTGTGGACCGACCGGCTCACCGCCGC
 CGACGTCTACCGCGCCAAGGCCTACAAGGTGGAGCCGGTGCCGGGCCAGGAAGGCCAGTAT
 TTCTGCTACATCGCCTATGATCTCGATTTGTTTCGAGGAAGGCTCCATCGCCAACCTCACGGCG
 10 TCGATCATCGGCAACGTCTTCTCCTTCAAGCCGCTGAAGGCGGCGCGGCTGGAGGACATGCG
 GCTTCCCGTCGCCTATGTGAAGACCTTCCGCGGCCCCGCCACCGGCATCGTGGTCGAGCGCG
 AGCGCCTGGACAAGTTCGGCCGCCCCCTTCTGGGCGCCACCACCAAGCCGAAGCTTGGCCTC
 TCGGGCAAGAATTACGGCCGCGTGGTCTATGAGGCCCTCAAGGGCGGCCTCGACTTCGTGAA
 GGACGACGAGAACATCAACTCGCAGCCCTTCATGCACTGGCGCGATCGTTCTCTATTGCAT
 15 GGAGGCCGTCAACAAGGCCCAGGCCGAGACCGGCGAGGTGAAGGGGCACTATCTCAACATC
 ACCGCCGGGACCATGGAGGAGATGTACCGCCGCGCCGAGTTCGCCAAGGAACTGGGCTCCG
 TGGTGGTGATGGTGGATCTCATCATCGGCTGGACCGCCATCCAGTCCATGTCCAACTGGTGC
 CGCGAGAACGACATGATCCTGCACATGCACCGTGCGGGCCATGGCACCTACACGCGCCAGAA
 GAGCCACGGCGTCTCCTTCCGCGTCATCGCCAAGTGGCTGCGGCTCGCCGGCGTCGACCACC
 20 TGCACACCGGCACCGCCGTGGGCAAGCTGGAAGGCGACCCCATGACCGTGCAGGGCTTCTA
 CAATGTCTGCCGCGAGACGACGACGACGACGAGGACCTCACCCGCGGCCTGTTCTTCGAGCAGG
 ACTGGGGCGGCATCCGCAAGGTGATGCCGGTGGCCTCCGGCGGCATCCATGCGGGCCAGAT
 GCACCAGCTCATCGACCTGTTTCGGCGAGGACGTGGTGTCCAGTTCGGCGGCGGCACCATCG
 GCCACCCGGACGGCATCCAGGCCGGCGCCACCGCCAACCGCGTGGCGCTGGAAACCATGAT
 25 CCTCGCCCCGAACGAGGGCCGCGACATCAGGAACGAGGGCCCGGAAATCCTGGTGGAAGCC
 GCCAAATGGTGCCGTCCGCTGCGCGCGGCGCTCGATACCTGGGGCGAGGTGACCTTCAACTA
 CGCCTCCACCGACACGTCCGATTACGTGCCCACCGCGTCCGTGCGCTGA

SEQ ID NO: 3:

30 Amino acid sequence of Ribulose bisphosphate carboxylase large chain
 MGAEATVGQITDAKKRYAAGVLKYAQMGYWNGDYVPKDTDLLAVFRITPQAGVDPVEAAA
 GESSTATWTVVWTDRLTAADVYRAKAYKVEPVPGQEGQYFCYIAYDLDFEEGSIANLTASIIGN
 VFSFKPLKAARLEDMRLPVAYVKTRFGPPTGIVVERERLDKFGRPLLGATTKPKLGLSGKNYGRVV
 YEALKGGLDFVKDDENINSQPFMHWRDRFLYCM EAVNKAQAETGEVKGHYLNITAGTMEEMYRR
 35 AEFKELGSVVVMVDLIIGWTAIQSMSNWCRENDMILHMHRAHGHTYTRQKSHGVSFRVIAKW
 LRLAGVDHLHTGTAVGKLEGDPMTVQGFYNVCRETTTQQDLTRGLFFEQDWGGIRKVMPPV
 ASG

GIHAGQMHQLIDLFGEDVVLQFGGGTIGHPDGIQAGATANRVALETMILARNEGRDIRNEGPEIL
VEAAKWCRPLRAALDTWGEVTFNYASTDTSDYVPTASVA

SEQ ID NO: 4:

- 5 Nucleotide sequence of Ribulose biphosphate carboxylase small chain:
ATGCGCATCACCCAAGGCTCCTTCTCCTTCTGCCGGACCTCACCGACACGCAGATCAAGGCC
CAGGTGCAATATTGCCTGGACCAGGGCTGGGCGGTCTCGGTGGAGCACACCGACGATCCCCA
CCCGCGCAACACCTATTGGGAGATGTGGGGCCCCGCCCATGTTTCGATCTGCGCGACGCGGCC
GGCGTCTTCGGCGAGATCGAAGCCTGCCGGGGCCGCCAATCCCGAGCATTATGTGCGGGTGAA
10 CGCCTTCGATTCCAGCCGCGGATGGGAGACGATCCGCCTGTCCTTCATCGTTCAGCGGCCCA
CCGTGGAAGAGGGCTTCCGCCTCGACCGCACCGAAGGCAAGGGCCGCAACCAGAGCTACGC
CATGCGCTACCGGGCGCAGTTCGCGCCGCGCTGA

SEQ ID NO: 5:

- 15 Amino acid sequence of Ribulose biphosphate carboxylase small chain:
MRITQGSFSFLPDLTDTQIKAQVQYCLDQGWAVSVEHTDDPHPRNTYWEMWGPPMFDLRDAAG
VFGEIEACRAANPEHYVRVNAFDSSRGWETIRLSFIVQRPTVEEGFRLDRTEGKGRNQSYAMRYR
AQFAPR
- 20 In another embodiment, the bacterial strain used in the process of the invention
comprises a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit alpha having
the sequence set forth in SEQ ID NO:7 or a sequence having more than 70% sequence
identity, such as more than 80% identity, e.g. more than 90% identity, such as more
than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g.
25 more than 98%, such as more than 99% sequence identity to the sequence set forth in
SEQ ID NO:7.

- In another embodiment, the bacterial strain used in the process of the invention
comprises a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit beta having the
30 sequence set forth in SEQ ID NO:9 or a sequence having more than 77% sequence
identity, such as more than 80% identity, e.g. more than 90% identity, such as more
than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g.
more than 98%, such as more than 99% sequence identity to the sequence set forth in
SEQ ID NO:9.

35

In another embodiment, the bacterial strain used in the process of the invention
comprises a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit gamma having

the sequence set forth in SEQ ID NO:11 or a sequence having more than 70% sequence identity, such as more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in
 5 SEQ ID NO:11.

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit delta having the sequence set forth in SEQ ID NO:13 or a sequence having more than 79% sequence
 10 identity, such as more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:13.

15 SEQ ID NO:6:

Nucleotide sequence of NAD⁺-reducing hydrogenase HoxS subunit alpha:

ATGATGCCATCTGAGCCGCACGGCGCGGGCATGCCGCCCCACGGGAAGCGGCCGCGGTTC
 CCACCCCCAGGAGGTGAGCGCGGTGGTGGCCGAGGTGGTCGCGGATGCCGTGGCATCGGT
 GGGCGGCGCACGCACCCGGCTCATGGACATCGTCCAGCTGGCCCAGCAGCGTCTCGGCCAT
 20 CTCTCCGAAGAGACCATGGCGGCCATTGCCGCGCGGCTCGCCATTCCGCCGGTGGAAGTGG
 CGGACATGGTGTCTTCTACGCCTTCCTGAACCGCGCGCCCAAGGGCCGCTACCACATCCGC
 CTGTCGCGCAGCCCCATCTCGTGATGAAGGGCGCCGAGGCGGTGGCTGCCGCCTTCTGCCA
 GATCCTCGGCATCGCCATGGGCGAGACCTCGCAGGATGGCGACTTCACCCTGGAATGGACCA
 ACGACATCGGCATGGCCGACCAGGAGCCGGCCGCCCTCGTCAACGGCACGGTGATGACGCA
 25 GCTCGCGCCCGGCGATGCGGCCATCATCGTCGGCCGGCTGCGGGCCCATCACGCGCCCAAT
 GCCCTGCCGCTGTTCCCTGGAGCCGGCGTGCGCGGCTCCGGCCTGCCCCATGCCCGGATCC
 GCCCCAGCCTGGTGATGCCGGGACAGCTTCTGTTCCGCGAGGACCACACGACGCCGGGCGC
 CGGCATCAAGGCGGCACTCGCCCTCACCCCGGACGAAGTGGTGAGAAGGTCTCCGCCGCG
 CGCCTGCGCGGGCGGGGTGGCGCCGGCTTTCCACCGGTCTCAAATGGAAGTCTGCGCGCC
 30 AGTCGCCCCGCCACCACCCGCCATGTGATCTGCAATGCGGACGAGGGCGAGCCCGGCACCTTC
 AAGGATCGCGTGCTGCTCACGCAGGCGCCGCACCTCATGTTTCGACGGCATGACCATCGCCGG
 CTACGCCTTGGGGGCGCGGGAGGGCGTGGTCTATCTGCGCGGCGAGTACGCCTATCTGTGG
 GAGCCTCTGCATGCGGTCCTGCGCGAGCGCTATGGGCTCGGGCTCGCCGGCGCGAACATCC
 TGGGACACGCGGGCTTCGACTTCGACATCCGCATCCAGCTGGGCGCCGGCGCCTATATCTGC
 35 GGCGAGGAATCCGCGCTGGTGGGAATCGCTGGAAGGCAAGCGCGGCTCGCCCCGCGACCGCC
 CCCCCTTCCCCACCGTGCGCGGCCATCTCCAGCAGCCCACCGCCGTGGACAATGTGGAGACC
 TTCGCCTGCGCCGCCCGCATCCTGGAGGATGGCGTGGAGGCGTTTCGCGGGCATCGGCACGC

CCGAATCCGCCGGCACGAAGCTCCTCTCGGTGTCGGGCGATTGCCCGCGCCCCGGCGTGTAT
 GAGGTGCCCTTCGGCCTCACGGTGAACGCGCTGCTCGACCTTGTGGCGCGCCGGACGCCG
 CCTTCGTGCAGATGGGTGGGCCGTCCGGCCAATGCGTGGCGCCGAAGGATTACGGCCGCCG
 CATCGCCTTCGAGGACCTGCCCACCGGCGGCTCGGTGATGGTGTTCGGCCCCGGGGCGCGAC
 5 GTGCTCGCCATGGTGCGCGAGTTCGCGGATTTCTTCGCCGGCGAATCCTGCGGCTGGTGAC
 GCCCTGCCGGGTGGGCACCACCTTGCTCAAGGAAGAGCTGGACAAGCTCCTCGCCAACCGCG
 CCACCCTCGCCGACATCCGCGCGCTGGAGACCCTGGCCACGACCGTCTCCCGCACCAGCCGC
 TGGCGCCTCGGCCAGACGGCGCCCAACCCCATCCTTTCCACCATGCGCAACCTGCCGGAAGC
 CTATGAGGCGAGGCTGAGGCCCCGAAGACTTCCTGCCCTGGGCCTCGCTCGACGAGGCGCTG
 10 AAGCCCGCCATCGTCATCCAGGGCCGCGCGCCCGTGCCGGAGGAAGAGGCATGA

SEQ ID NO: 7:

Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit alpha:

MMPSEPHGAGMPPPREAAVPTPQEVSAVVAEVDADAVASVGGARTRLMDIVQLAQQLGHLSE
 15 ETMAAIAARLAIPPVEVADMVSFYAFLNRAPKGRYHIRLSRSPISLMKGAEAVAAAFQCILGIAMG
 ETSQDGDFTLEWTNDIGMADQEPAALVNGTVMTQLAPGDAAIIVGRLRAHHAPNALPLFPAGV
 AGSGLPHARIRPSLVMPGQLLFREDHTTPGAGIKAALALTPDEVVQKVSAARLRGRGGAGFPTGL
 KWKLCRQSPATTRHVICNADEGEPGTFKDRVLLTQAPHLMFDGMTIAGYALGAREGVVYLRGEY
 AYLWEPLHAVLRERYGLGLAGANILGHAGFDFFIRIQLGAGAYICGEESALVESLEGKRGSPRDRP
 20 PFPTVRGHLQQPTAVDNVETFACAAIRILEDGVEAFAGIGTPESAGTKLLSVSGDCPRPGVYEVFPF
 LTVNALLDLVGAPDAAFVQMGGPSGQCVAPKDYGRRIAFEDLPTGGSVMVFGPGRDVLAMVREF
 ADFFAGESCGWCTPCRVTLLKEELDKLLANRATLADIRALETLATTVSRTSRCGLGQTAPNPIL
 STMRNLPEAYEARLRPEDFLPWASLDEALKPAIVIQGRAPVPEEEA

25 SEQ ID NO: 8:

Nucleotide sequence of NAD⁺-reducing hydrogenase HoxS subunit beta:

ATGAGCCGGGGATCCCCCGATGCCGGGAAAGACCGCACCATGAGCGCCACCGACGGCACCA
 CCGCCCCCGCAAGATCGTCATCGATCCGGTGACCCGCGTGAGGGGCCACGGCAAGGTCAC
 CATCCGCCTGGATGAAGCCGGCGCGGTGGAGGATGCGCGTTTCCACATCGTGGAGTTCCGC
 30 GGCTTCGAGCGGTTTCATCCAGGGCCGGATGTACTGGGAAGTGCCCCTTATCATCCAGCGGCT
 GTGCGGCATCTGCCCGGTGAGCCACCATCTGGCGGCGGCGAAAGCCATGGACCAGGTGGCG
 GCGTGACCGCGTACCGCCCACCGCCGAGAACTGCGCCGGCTGATGCATTATGGGCAGG
 TGCTGCAATCCAACGCTTTGCACATCTTCCACCTCGCCTCGCCGACCTCCTGTTCCGGCTTCG
 ACGCGCCGGCCGAGCAGCGCAACATCATCGCCGTGCTCCAGCGTTATCCGGAGATCGGCAAA
 35 TGGGCGATCTTCATCAGGAAGTTCGGCCAGGAGGTCATCAAGGCCACCGGCGGGCGCAAGA
 TCCATCCCACCAGCGCCATTCCCGGCGGGGTCAACCAGAACCTCGCCGTGGAGGACCGCGAC
 GCCCTGCGCGCCAAGGTGGGCGAGATCATCAGCTGGTGCATGGCGGCGCTGGACCATCACA

AGGCCTATGTGGCGGAAAACCGGGCGCTGCATGACAGCTTCGCCGCCTTCCCCTCCGCCTTC
 ATGAGCCTCGTGGGGCCGGATGGCGGCATGGACCTTTATGACGGCACCTGCGGGTGATCG
 ATGCCGAGGGCGCCCCCTCATCGAAGGCGCGCCGCCGCCTCCTACCGCGACCACCTCATC
 GAGGAGGTGCGGCCCTGGAGCTATCTGAAATTCCCCATCTGCGCGCCTTCGGCCGCGACGA
 5 TGGCTGGTATCGGGTCGGCCCCCTCGCCCAGGTCAATTGCGCCGCGTCCATCGACACGCCCC
 GCGCCGAGGCGGCCCGGCGGGACTTCATGGCCGAGGGCGGCGGCAAGCCGGTGCATGCCA
 CCCTCGCTTATCACTGGGCGCGGCTCATCGTGCTGGTCCATTGCGCGGAGAAGATCGAACAG
 CTGCTGTTTCGACGACGACCTGCAAGGCTGCGATCTGCGTGCGGAGGGCACCCGGCGCGGGG
 AAGGCGTCGCCTGGATCGAGGCGCCGCGCGGCACCCTCATCCACCATTACGAGGTGGACGA
 10 GAACGACCAGGTGCGCCGCGCCAACCTCATCGTCTCCACCACCCACAATAACGAGGCCATGA
 ACCGCGCCGTGCGGCAGGTGGCGAAGACGGACCTTTCCGGTCGCGAGATCACCGAAGGGCT
 GCTGAACCATATCGAGGTGGCCATCCGCGCCTTCGACCCCTGCCTGTCCTGCGCCACCCATG
 CGCTGGGCCAGATGCCGCTGATCGTGACGCTTGAAGATGCCTCCGGCGCAGAGATCGCCCG
 CGGAGTGAAGGAATGA

15

SEQ ID NO: 9:

Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit beta:

MSRGSPDAGKDRTMSATDGTAPRKIVIDPVTRVEGHGKVTIRLDEAGAVEDARFHIVEFRGFER
 FIQGRMYWEVPLIIQRLCGICPVSHHLAAAKAMDQVAGVDRVPPTAEKLRLMHYGQVLQSNAL
 20 HIFHLASPDLLFGFDAPAEQRNIIAVLQRYPEIGKWAIFIRKFGQEVIKATGGRKIHPTSAIPGGVN
 QNLAVEDRDALRAKVGEIISWCMAALDHHKAYVAENRALHDSFAAFPSAFMSLVGPDGGMDLY
 DGTLRVIDAEGAPLIEGAPPASYRDHLIEEVRPWSYLFPHLRAFGRDDGWYRVGPLAQVNCAAS
 IDTPRAEAARRDFMAEGGGKPVHATLAYHWARLIVLVHCAEKIEQLLFDDDLQGCDLRAEGTRRG
 EGVAWIEAPRGTLIHHEVDENDQVRANLIVSTTHNNEAMNRAVRQVAKTDLSGREITEGLLN
 25 HIEVAIRAFDPCLSCATHALGQMPLIVTLEDASGAEIARGVKE

SEQ ID NO: 10:

Nucleotide sequence of NAD⁺-reducing hydrogenase HoxS subunit gamma:

ATGAGCGAGACCCCTTCACCTTTACCGTGGACGGCATCGCGGTCCCGGCCACCCCGGCCA
 30 GAGCGTCATCGAGGCGTGCGATGCGGCGGGCATCTATATCCCGCGCCTGTGCCACCACCCG
 GACCTGCCGCCGGCGGGCCATTGCCGGGTGTGCACCTGCATCATCGACGGGCGGCCGGCCA
 GCGCCTGCACCATGCCCCGCCGCGAGGGGCATGGTGGTGGAGAACGAGACGCCCGCTTTGCT
 GGCGGAGCGGCGCACGCTGATCGAGATGCTGTTTCGCGGAAGGCAACCATTTCTGCCAGTTCT
 GCGAGGCGAGCGGCGATTGCGAATTGCAGGCGCTGGGCTACCTGTTTCGGCATGGTGGCCCC
 35 GCCCTTCCCCCATCTGTGGCCGAAGCGGCCGGTGGATGCCAGCCATCCGGATATCTATATCG
 ACCACAATCGCTGCATCCTGTGCTCGCGCTGCGTGCGCGCCTCGCGCACCTGGACGGCAAG
 TCCGTGTTTCGGCTTCGAGGGGCGCGGCATCGAGATGCATCTGGCGGTGACCGGCGGGCACC

TGGACGACAGCGCCATCGCCGCCGCCGACAGGGCGGTTGAGATGTGCCCCGGTGGGCTGCAT
CGTCTCAAGCGCACCGGCTACCGCACGCCCTATGGCCGGCGGCGCTACGACGCCGCGCCC
ATCGGCTCCGACATCACCGCCCCGGCGCGGCGGCGCGAAGGACTGA

5 SEQ ID NO: 11:

Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit gamma:

MSETPFTFTVDGIAVPATPGQSVIEACDAAGIYIPRLCHHPDLPPAGHCRVCTCIIDGRPASACTM
PAARGMVVENETPALLAERRTLIEMLFAEGNHFCQFCEASGDCELQALGYLFGMVAPPPHLWPK
RPVDASHPDIYIDHNRCILSRCVRASRTLDGKSVFGFEGRGIEMHLAVTGGHLDDSAIAAADRA
10 VEMCPVGCIVLKRTGYRTPYGRRRYDAAPIGSDITARRGGAKD

SEQ ID NO: 12:

Nucleotide sequence of NAD⁺-reducing hydrogenase HoxS subunit delta:

ATGGCCAAGCCCAAACTCGCCACCTGCGCGCTGGCCGGCTGCTTCGGCTGCCACATGTCCTT
15 CCTGGACATGGACGAGCGCATCGTCGAGCTCATCGACCTGGTGGACCTCGACGTCTCGCCCC
TCGACGACAAGAAAACTTCACCGGCATGGTGGAAATCGGCCTGGTGGAAAGGCGGCTGCGC
CGACGAGCGCCATGTGAAGGTGCTGCGCGAGTTCCGCGAGAAATCCCGCATCCTGGTGGCG
GTGGGCGCCTGCGCCATCACCGGCGGCATCCCGGCATTGCGCAACCTCGCCGGCCTCGACG
AATGCCTGAGGGAAGCCTACCTACCGGCCCCACGGTGGAAAGGCGGCGGGCTCATTCCCAAC
20 GACCCGAGCTGCCGCTGCTGCTGGACAAGGTCTATCCGGTGCAGGACTTCGTGAAGATCGA
CCATTTCTGCCCGGCTGCCCGCCCTCGGCCGACGCCATCTGGGCGGCTCTGAAGGCGCTGC
TGACCGGCACCGAGCCGCATCTGCCCTACCCGCTTTTCAAGTACGAATGA

SEQ ID NO: 13:

25 Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit delta:

MAKPKLATCALAGCFGCHMSFLDMDERIVELIDLVDLDVSPLDDKKNFTGMVEIGLVEGGCADER
HVKVLREFREKSRLVAVGACAITGGIPALRNLAGLDECLREAYLTGPTVEGGGLIPNDPELPLLLD
KVYPVQDFVKIDHFLPGCPPSADAIWAALKALLTGTEPHLPYPLFKYE

30 In another embodiment, the bacterial strain used in the process of the invention
comprises a gene encoding a NiFeSe hydrogenase large subunit having the sequence
set forth in SEQ ID NO:15 or a sequence having more than 84% sequence identity, e.g.
more than 90% identity, such as more than 95% identity, e.g. more than 96% identity,
such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence
35 identity to the sequence set forth in SEQ ID NO:15.

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding a NiFeSe hydrogenase small subunit having the sequence set forth in SEQ ID NO:17 or a sequence having more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:17.

SEQ ID NO: 14:

Nucleotide sequence of Periplasmic [NiFeSe] hydrogenase large subunit:

10 TCCAGACCCGGGCAACATTGCTCCATGTGCTGGGCACCTGGCCGGCCGCTGGCCCCATACC
CTCGCGCTCCAGCCCGGCGGGGTGACCCGAAGCGCCGACCAGCACGACCGCATGCGCCTGC
TCGCGACGCTGAAGGCGGTGCGGGCGGCGCTGGAAGAGACCTTGTTGCGCGCGCCTTTGGA
AGAGGTGGCGGCCCTGGACGGCGCCGCCGCGTGGAGGCCTGGCGCGCCAACGGCCCCGGA
AGGGGATTTCCGCCTGTTCTGAGATCGCCGCCGACCTGGAGCTGGACCGGCTCGGCCGC
15 GCGCACGACCGCTTTCTCTCCTTCGGCGCCTACGCCAGGACGAGGGGCGCCTTTATGGCGC
CGGCACCTTCGAGGCCGGGACGGCGGGAGGGCTCGATCCCAACGCCATCACCGAGGACCAC
GCCTTCGCCCCGCATGGAGGACCGCGCGGGCGCCCCATGCGCCCTTTGACGGCTCCACCTCCC
CGATGCCGACGACACCGAGGGCTACACCTGGTGCAAGGCGCCGCGCCTTGCCGGCCTGCCC
TTCGAGACCGGCGCCTTCGCCCCGCGAGGTGGTGGCGGGCCATCCGCTCGCCCGGGACCTCG
20 TGACGCGGGAAGGCGGCACTGTGCGCAGCCGCGTGGTCGGCCGGCTGCTGGAAACCGCGC
GCACCCTGATCGCCATGGAGGGCTGGGTGAAGGAACTGCGGCCCCGAAGGGCCCTGGTGCGC
CCAGGGCCACCTGCCCCAGGAAGGCCGCGCCTTCGGCCTCACCGAGGCGGCGCGCGGGGC
GCTCGGCCACTGGATGGTGGTGGAGAAGGGCCGCATTGCCCGCTACCAGATCATCGCCCCCA
CCACCTGGAACCTTCTCCCCCGCGACGGCGCGGGCCTGCCCGGCCCGCTGGAGACGGCCCT
25 GGTGGGCGCGCCCGTGCGGCAGGGAGAGACGACGCCCGTGAGCGTGCAGCACATCGTGCG
CTCCTTCGACCCGTGCATGGTCTGCACTGTGCATTGA

SEQ ID NO: 15:

Amino acid sequence of Periplasmic [NiFeSe] hydrogenase large subunit:

30 MSAETRRLVVGPFNRVEGDLEVRDLVDGRVQQAFVSSPLFRGFERILEGRDPRDALVIAPRICGI
CSVSQSHAAALALAGLQGIAPTHDGRIATNLIVAAENVADHLTHFHVFFMPDFARAVYEDRPWFA
QAARRFKANQGVSVRRALQTRATLLHVLGTLAGRWPHTLALQPGGVTRSADQHDMRLLATLKA
VRAALEETLFGAPLEEVAAALDGAAAVEAWRANGPEGDFRLFLEIAADLELDRLGRAHDRFLSFGAY
AQDEGRLYGAGTFEAGTAGGLDPNAITEDHAFARMEDRAAPHAPFDGSTFPDADDTEGYTWCK
35 APRLAGLPFETGAFARQVVAGHPLARDLVTREGGTVRSRVVGRLLLETARTLIAMEGWVKELRPEG
PWCAQGHLPQEGRAFGLTEAARGALGHWMVVEKGRIARYQIIAPTTWNFSPRDGAGLPGPLETA
LVGAPVRQGETTPVSVQHIVRSFDCMVCTVH

SEQ ID NO: 16:

Nucleotide sequence of Periplasmic [NiFeSe] hydrogenase small subunit

ACGGGGGAGGAAGCCCGCGCCATCTTCGACGCCATCCTTGCCGGCGTTATCGTCCTCGACGC
 5 CCTGTGCGTGGAAGGCGCGCTGCTGCGCGGGCCGAACGGCACCGGGCGCTTCCATGTGCTG
 GCGGGCACGGACACCCCCACCATCGACTGGGCGCGGCAGCTCGCCGGCATGGCGCGCCACG
 TGGTGGCGGTGGGCACCTGCGCCGCCTATGGGGGCGTGACGGCGGCGGGCATCAACCCAC
 CGATGCCTGCGGCCTCCAGTTCGACGGACGCCGGAAGGGTGGGGCGCTGGGGGCGGACTTC
 CGCTCCCGCTCGGGGCTTCCGGTCATCAATGTGGCCGGCTGCCCCACCCATCCCAACTGGGT
 10 GACGGAAACCCTGATGCTGCTCGCCTGCGGCCTGCTGGGCGAGGCCGACCTCGACGTCTATG
 GCCGCCCCGCGCTTCTATGCGGACCTGCTGGTGCATCACGGCTGCCGCGCAACGAATACTAT
 GAATACAAGGCGAGCGCCGAGAAGATGAGCGACCTCGGCTGCATGATGGAGCATCTGGGCT
 GCCTCGGCACCCAGGCCACGCCGACTGCAACACGCGCCTTTGGAATGGCGAGGGCTCGTG
 CACCCGCGGCGGCTATGCCTGCATCAACTGCACGGCGCCGGAATTCGAGGAGCCGGGCCAC
 15 GCCTTCCTGGAGACGCCCAAGATCGGCGGCATCCCCATCGGCCTGCCACCGACATGCCCAA
 GGCCTGGTTCATCGCCTTGCTCCTCCCTCGCCAAGGCGGCGACGCCGGAGCGGCTGCGCAAG
 AACGCGGTGTCCGACCATGTGGTCACGCCGCCCGCCGTCAAGGACATCAAGCGGCGATGA

SEQ ID NO: 17:

Amino acid sequence of Periplasmic [NiFeSe] hydrogenase small subunit
 20 MSTPFSVLWLQSGGCGGCTMSLLCAEAPDLATTLDAAGIGFLWHPALSEETGEEARAIFDAILAG
 VIVLDALCVGALLRGPNGTGRFHVLAGTDTPTIDWARQLAGMARHVVAVGTCAAYGGVTAAGI
 NPTDACGLQFDGRRKGGALGADFRSRGLPVINVAGCPHPNWWVTETLMLLACGLLGEADLDVY
 GRPRFYADLLVHHGCPRNEYEYKASAEKMSDLGCMMEHLGCLGTQAHADCNTRLWNGEGSCT
 25 RGGYACINCTAPEFEPEGHAFLETPKIGGIPIGLPTDMPKAWFIALSSLAKAATPERLRKNAVSDHV
 VTPPAVKDIKRR

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase gamma chain atpG_1 having the sequence set forth in SEQ ID NO:19 or a sequence having more than 70% identity, such as more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:19.

35 In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit alpha atpA_1 having the sequence set forth in SEQ ID NO:21 or a sequence having more than 78% identity, such as more

than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:21.

- 5 In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit b atpF_1 having the sequence set forth in SEQ ID NO:23 or a sequence having more than 62% identity, e.g. more than 70% identity, such as more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, 10 e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:23.

- In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit c, sodium ion specific atpE_1 15 having the sequence set forth in SEQ ID NO:25 or a sequence having more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:25.

- 20 In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit a atpB_1 having the sequence set forth in SEQ ID NO:27 or a sequence having more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity 25 to the sequence set forth in SEQ ID NO:27.

- In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase epsilon chain atpC_1 having the sequence set forth in SEQ ID NO:29 or a sequence having more than 71% identity, such as more 30 than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:29.

- In another embodiment, the bacterial strain used in the process of the invention 35 comprises a gene encoding an ATP synthase subunit beta atpD_1 having the sequence set forth in SEQ ID NO:31 or a sequence having more than 84% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such

as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:31.

5 In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit beta atpD_2 having the sequence set forth in SEQ ID NO:33 or a sequence having more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:33.

10 In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase gamma chain atpG_2 having the sequence set forth in SEQ ID NO:35 or a sequence having more than 86% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence
15 identity to the sequence set forth in SEQ ID NO:35.

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit alpha atpA_2 having the sequence set forth in SEQ ID NO:37 or a sequence having more than 98%, such as more than
20 99% sequence identity to the sequence set forth in SEQ ID NO:37.

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit delta atpH having the sequence set forth in SEQ ID NO:39 or a sequence having more than 85% identity, e.g. more
25 than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:39.

In another embodiment, the bacterial strain used in the process of the invention
30 comprises a gene encoding an ATP synthase subunit b atpF_2 having the sequence set forth in SEQ ID NO:41 or a sequence having more than 87% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:41.

35

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit b' atpG_3 having the sequence set

forth in SEQ ID NO:43 or a sequence having more than 81% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:43.

5

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding ATP synthase subunit c atpE_2 having the sequence set forth in SEQ ID NO:45 or a sequence having more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:45.

10

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase subunit a atpB_2 having the sequence set forth in SEQ ID NO:47 or a sequence having more than 92% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:47.

15

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding an ATP synthase protein I atpI having the sequence set forth in SEQ ID NO:49 or a sequence having more than 60% identity, e.g. more than 70% identity, such as more than 80% identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:49.

25

SEQ ID NO: 18:

Nucleotide sequence of ATP synthase gamma chain atpG_1

GTGACCGAGCGCCTGTCCGACGTCAACGCCCGCATCGCCTCGGTGCGGCAGCTCTCATCGGT
 CATCACGGCCATGCGGGGCATTGCGGCGGCGCGGGCGCGGGAGGCGCGGGGTCGGCTCGA
 30 CGGCATCCGCGCCTATGCGCAGACCATCGCCGAGGCCATCGGCCATGTGCTCGCCGTGCTGC
 CCGAGGAGGCCCCGCGCCCGGTCTCCGGGCACCGGCATCGGGGCCATGCGGTCATCGCCCT
 GTGCGCGGAGCAGGGCTTTGCCGGCGTCTTCAACGAGCGGGTGCTGGACGAGGCCGCCCGG
 CTGCTGACCGGCGGGGCGGGGCCGGCCGAGCTGCTGCTGGTGCGGACCGGGGCCTGATG
 GTGGCCCGCGAGCGGGGGCTCGATGTCTCCTGGTCGGTGCCCATGGTGGCCCATGCGGGCC
 35 AGGCCTCGGCGCTGGCGGACCGCATCAGCGAGGAGCTCTACCGGCGGATCGATGCGGGACG
 GGTGACGCGGGTGTCGGTGGTGCACGCCGAGCCCGCCGCGTCCGCCGCCATCGAGACGGTG
 GTGAAAGTGCTGGTGCCGTTGACTTCGCCCGCTTCCCCCTGGCGCGGGTGGCATCCGCCCC

GCTCATGACCATGCCGCCGCCGCGGCTGCTGGCCCAGCTGTCGGAGGAATATGTGTTGCGCCG
 AGCTGTGCGAGGCGCTCACCTTGTCTTCGCGGCGGAGAACGAGGCCCGCATGCGGGCCAT
 GATCGCCGCCCGCGCCAATGTGGCCGATACCCTGGAGGGCCTCGTCGGCCGCGCCCGGCAG
 ATGCGCCAGGAGGAGATCACCAACGAGATCATCGAGCTGGAAGGCGGCGCCGGCAGCGCCC
 5 GGCATGCGGATTGA

SEQ ID NO: 19:

Amino acid sequence of ATP synthase gamma chain atpG_1

MTERLSDVNARIASVRQLSSVITAMRGIAAARAREARGRLDGIRAYAQTIAEAIGHVLAVLPPEAR
 10 ARSSGHRHRGHAVIALCAEQGFAGVFNERVLDEAARLLTGGAGPAELLVGDRLMVARERGLD
 VSWSVPMVAHAGQASALADRISEELYRRIDAGRVTRVSVVHAEPAAASAAIETVVKVLVPFDFARF
 PLARVASAPLMTMPPPRLLAQLSEEYVFAELCEALTLSFAAENEARMRAMIAARANVADTLEGLVG
 RARQMRQEITNEIIELEGGAGSARHAD

15 SEQ ID NO: 20:

Nucleotide sequence of ATP synthase subunit alpha atpA_1

ATGAGCACGGGCGCGCAAGCGAGCGAGGATTGGCTACCCGGAGCCGGGCGGCCCTGGCC
 GGGACGCGCCTTTCCAGCAATCCCAATCGGTGGGCGGGTGGAGGAGATGGCCGACGGCA
 TCGCCCGCGTCTCCGGCCTGCCGGATGTGCGGCTCGACGAGCTTCTCACCTTCGAGGGCGGC
 20 CAGACCGGCTATGCCCTCACCTCGATCGCACCGAGATCGCCGTGGTGTCTGCTGGATGACGC
 CTCCGGCGTGGAGGCGGGCGCCCGGGTGTTCGGCACCGGCGAGGTGGTGAAGGTGCCGGT
 GGGGCCGGGGCTGCTGGGCCGCATCGTCGACCCCTCGGCCGGGCCATGGACCGCTCCGAG
 CCGGTGGTGGCGCAGGCGCACCATCCCATCGAGCGGCCGGCGCCGGCCATCATCGCCCGCG
 ACCTGGTCTCGCAGCCGGTTCAGACCGGCACGCTGGTGGTGGATGCGCTGTTCTCCCTCGGC
 25 CGGGGCCAGCGCGAGCTCATCATCGGCGACCGGGCTACCGGCAAGACCGCCATCGCGGTGG
 ACACCATCATCAGCCAGAAGCATTCGGACATCGTGTGCATCTACGTGGCGGTGGGCCAGCGC
 GCCGCCGCCGTGGAGCGGGTGGTGGAGGCGGTGCGCGCCACGGGGCGATCGAGCGCTGC
 ATCTTCGTGGTGCCTCGGCCGCCCTCGCCAGGGCTGCAATGGATCGCGCCGTTCCGCCG
 CATGACCATGGCGGAATATTTCCGCGACAACGGCCAGCATGCGCTCATCATCATGATGATCT
 30 CACCAAGCATGCGGCCACCCATCGCGAGCTGGCGCTGCTACCCACGAGCCGCCGGGCCGC
 GAGGCCTATCCCGGCGACATCTTCTATGTGCACGCCCGCCTTCTGGAGCGGGCCGCCAAGCT
 CTCCGCCGAGCTGGGCGGTGGCTCGCTCACGGCCCTGCCCATCGCGGAGACGGACGCGGGA
 AACCTCTCCGCCTATATCCCCACCAACCTCATCTCCATCACCGATGGGCAGATCGTGCTGGAT
 TCGCGGCTGTTTCGCGGCCAACCAGCGCCCGGCGGTGGATGTGGGCCTCTCCGTGAGCCGGG
 35 TGGGCGGCAAGGCGCAGCATCCCGCGCTTCGGGCCGTGTCCGGGCGCATCCGGCTCGATTA
 TTCCCAGTTCCTGGAGCTGGAAATGTTACCCGCTTCGGCGGCATCACCGATACCCGCGTGAA
 GGCGCAGATCACCCGGGGCGAGCGCATCCGCGCGCTGCTACCCAGCCGCGCTTTTCCACCC

TGCGCCTTCAGGACGAGGTGGCGCTGCTGGCCGCGCTGGCGGAGGGGGTGTTCGACACTTT
 GGCCCCGGGGCTGATGGGCGCCGTGCGTGCCCGCATTCCGGCCCAGCTGGATGCGCAGGTG
 AAGGACGTGGCCTCGGCCCTCGCCGAGGGCAAGGTGCTGGAGGAGGGCTTGCACGCCCCGTC
 TCGTGGCGGCCGTGCGGGCCGTGCGGGCGGACGTGGCCGCGACCGCGAAGGCCGGGCCGT
 5 GA

SEQ ID NO: 21:

Amino acid sequence of ATP synthase subunit alpha atpA_1

MSTGAQASEDWLTRSRAALAGTRLSQQSQSVGRVEEMADGIARVSGLPDVRLDELLTFEGGQT
 10 GYALTLDRTEIAVVLLDDASGVEAGARVFGTGEVVKVPVGPGLLGRIVDPLGRPMRSEPVVAQA
 HHPIERPAPAIARDLVSQPVQTGTLVVDALFSLGRGQRELIIGDRATGKTAIAVDTIISQKHSDIV
 CIYVAVGQRAAAVERVVEAVRAHGAIERCIFVVASAAASPLQWIAPFAGMTMAEYFRDNGQHA
 LIIIDDLTKHAATHRELALLTHEPPGREAYPGDIFYVHARLLERAALKSAELGGGSLTALPIAETDAG
 NLSAYIPTNLISITDGQIVLDSRLFAANQRPADVGLSVSRVGGKAQHPALRAVSGRIRLDYSQFL
 15 ELEMFTFRGGITDTRVKAQITRGERIRALLTQPRFSTLRLQDEVALLAALAEGVFDLAPGLMGAV
 RARIPAQLDAQVKDVASALAEQKVLLEGLHARLVAAVRAVAADVAATAKAGP

SEQ ID NO: 22:

Nucleotide sequence of ATP synthase subunit b atpF_1

20 ATGCAGATCGACTGGTGGACGCTGGGCCTGCAGACGGTCAACGTCCTCGTTCTCATCTGGCT
 CCTGAGCCGCTTCCTGTTCAAGCCGGTGGCGCAGGTCATCGCGCAGCGCCGTGCCGAGATCG
 AGAAGCTGGTGGAGGATGCGCGCGCCGCCAAGGCCGCCGCCGAGGCCGAGCGGGACACGG
 CGAAGGCGGAGGAGGCGCGCCTTGCCGCCGAGCGCGGCGCCCGCATGGCGGCGGTGCGCA
 AGGAGGCGGAGGCGCAGAAGGCGGCATTGCTGGCCGCCGCCAAGACCGAGGCCGAGGCC
 25 TGCACGCGGCCGCGGAAGCGGCCATCGTCCGGGCGCGGGCGAGCGAGGAGGAAGCCGCCG
 CCGACCGCGCCAGCCGCTTGCCGTGGACATCGCCGCCAAGCTGCTGGACCGGCTGCCCGA
 CGACGCCCCGGGTCGCGGGCTTCATCGATGGCCTCGCCGAGGGGCTTGAAGCCCTGCCCGAG
 GCGAGCCGGGCGGTGATCGGCGTCGACGGCGCGCCAGTGC GCGTGACGGCCGCGCGCGCC
 CTTATGCCGGCGGAGGAGGAGGCCTGCCGCACGCGGCTCTCCCAGGCGCTGGGCGCTCCGG
 30 TGACGCTGGCCGTGACCATCGACCCGCCCTCATCGCCGGCCTGGAGATGGAGACGCCCCAC
 GCGGTGGTGC GCAATTCCTTCAAGGCCGATCTCGACCGCGTCACCGCGGCGCTCACCCATCA
 TGGGACCTGA

SEQ ID NO: 23:

35 Amino acid sequence of ATP synthase subunit b atpF_1

MQIDWWTLGLQTVNVLVLIIWLLSRFLFKPVAQVIAQRRAEIEKLVEDARAAKAAAEAEERDTAKAE
 EARLAAERGARMAAVAKEAEQAALLAAKTEAEALHAAAEAAIVRARASEEEAAADRASRLAV

24

DIAAKLLDRLPDDARVAGFIDGLAEGLEALPEASRAVIGVDGAPVRVTAARALMPAEEEEACRTRLS
QALGRPVT LAVTIDPALIAGLEMETPHAVVRNSFKADLDRVTAALTHHGT

SEQ ID NO: 24:

5 Nucleotide sequence of ATP synthase subunit c, sodium ion specific atpE_1
ATGACTGTCGAGATGGTCAGCATCTTCGCGGCGGCGCTCGCCGTCTCCTTCGGCGCCATCGG
GCCGCCCCCTGGGCGAGGGCCGGGCGGTGGCCGCGGCCATGGACGCCATCGCCCGCCAGCC
GGAGGCGGCCGGAACCTTGTCGCGCACGCTCTTCGTCGGCCTCGCCATGATCGAGACCATGG
CGATCTACTGCCTGGTGATCGCGCTCCTGGTGCTCTTCGCCAATCCGTTCGTGAAGTGA

10

SEQ ID NO: 25:

Amino acid sequence of ATP synthase subunit c, sodium ion specific atpE_1
MTVEMVSIFAAALAVSFGAIGPALGEGRAVAAAMDAIARQPEAAGTLSRTL FVGLAM IETMAIYCL
VIALLVLFANPFVK

15

SEQ ID NO: 26:

Nucleotide sequence of ATP synthase subunit a atpB_1
ATGGGCTCGCCGCTGATCCTCGAACCCCTGTTCCATATCGGGCCCGTGCCCATCACCGCGCC
GGTGGTGGTCACCTGGCTCATCATGGCCGCCTTCATTGGGCTGGCGCGGCTCATCACCCGGA
20 AGCTTTCCACCGATCCCACCCGGACCCAGGCGGCGGTGGAAACGGTGCTGACCGCCATCGAT
TCCCAGATCGCCGACACCATGCAGGCCGATCCCGCGCCTTATCGCGCGCTCATCGGCACCAT
CTTCCTTTATGTGCTGGTGGCCAACTGGTCCTCGCTCATCCCGGGCATCGAGCCGCCCACGG
CGCATATCGAGACCGATGCGGCGCTCGCTTTCATCGTGTTGCGCGCCACCATCGGGTTCGGG
TTGAAGACAAGGGGTGTGAAGGGCTATCTCGCCACCTTCGCCGAACCCTCCTGGGTGATGAT
25 CCCGCTCAATGTGGTGGAGCAGATCACCCGGACCTTCTCGCTCATCGTGCGCCTGTTCGGCA
ACATCATGAGCGGGGTGTTCTGTGGTCGGCATCATCCTGTCCCTCGCCGGGCTGCTGGTGCCC
ATCCCCCTCATGGCGCTCGATCTCCTGACCGGCGCCGTGCAGGCCTACATCTTCGCGGTGCT
GGCCTGCGTGTTTCATCGGCGCGGCCATTGGCGAGGCGCCGGCAAAGCCCCAATCGAAGGAG
CCAGGGAAAACATCATGA

30

SEQ ID NO: 27:

Amino acid sequence of ATP synthase subunit a atpB_1
MGSPLILEPLFHIGVPITAPVVVTWLIMAAFIGLARLITRKLSTDPTRTQAAVETVLT AID SQIADT
MQADPAPYRALIGTIFLYVLVANWSSLIPGIEPPTAH IETDAALAFIVFAATIGFLKTRGVKGYLAT
35 FAEPSWVMIPLNVEQITRTFSLIVRLFNGIMSGVFVVGII LSLAGLLVPIPLMALDLLTGAVQAYIF
AVLACVFIGAAIGEAPAKPQSKEPGKTS

SEQ ID NO: 28:

Nucleotide sequence of ATP synthase epsilon chain atpC_1

GTGAGCGCGCCGCTGCACCTCACCATCACCACGCCGCCGCGTCTGGTGGACCGTGCCGA
 CATCGTGGCCCTGCGTGCCGAGGACGAGAGCGGCAGCTTCGGCATCCTGCCCCGCCATGCG
 5 GATTTCTGACCGTTCTGGAGGCCTGCGTGGTGCCTTCAAGGATGGGGCCGACGGCGTGCA
 TTATTGTGCTCTCAGTGGTGGCGTGCTGTGGTTCGAGGAGGGCCGGCGCATCGCCATCGCCT
 GCCGTGAGGGCACGGTGAGCGACGACCTGGTCGCCCTGGAAGGGGCGGTGGACGCCATGC
 GTTCGGCGGAGAGCGATGCCGACAAGCGGGCCCGGGTGGAGCAGATGCGCCTTCATGCCCA
 CGCCGTGCGCCAGCTCCTGCACTATCTGCGGCCCGGCCGGGCCGGCGGTGGCGCCGGCC
 10 GCCGCGCCGGAGGAGGGGCCGTCATGA

SEQ ID NO: 29:

Amino acid sequence of ATP synthase epsilon chain atpC_1

MSAPLHLTITTPAAVLVDRADIVALRAEDESFSFILPGHADFLTVLEACVVRFKDGADGVHYCAL
 15 SGGVLSVEEGRRIAIAICRQGTVSDDLVALEGAVDAMRSAESDADKRARVEQMRLHAHAVRQLL
 HYLPRGRAGGVAPAAPEEGPS

SEQ ID NO: 30:

Nucleotide sequence of ATP synthase subunit beta atpD_1

ATGGCAGCGGCAGATGAGGAGGCGCAATCGGCCGCCGCCCGCCTCGGGCCGGGTGGTG
 20 GCCGTGCGCGGCGCGGTGATCGACATCGCCTTTGCCAGCCTCCGCTGCCGCCGCTGGACG
 ACGCCCTTCTCATCACCGACGGCCGGGCGGCACGGTGCTGGTGGAGGTGCAGAGCCATAT
 GGATCGGCACACGGTGCGCGCCATCGCCCTTCAGGCCACCACCGGCCTCAGCCGGGGGCTG
 GAGGCGGCGCGGGTGGGCGGGCCGGTGAAGGTGCCGGTGGGAGACCATGTGCTCGGCCGC
 25 CTCCTGGATGTCACCGGCGCCATCGGCGACAAGGGCGGGCCGCTGCCGGCCGACGTGCCCA
 CGCGGCCGATCCACCACGCGCCGCCATCCTTCGCCGCGCAGGGCGGCACGTCCGATCTGTTT
 CGCACCGGCATCAAGGTCATCGACCTCCTGGCGCCCCCTCGCCCAGGGCGGCAAGGCGGCCA
 TGTTCCGGCGGGGCCGGCGTGGGCAAGACCGTGCTGGTGATGGAGCTGATCCACGCCATGGT
 GGCAGCTACAAGGGCATCTCGGTGTTTGCCGGCGTGGGGGAGCGCTCCCGCGAGGGCCAC
 30 GAGATGCTGCTGGACATGACCGATTCCGGCGTGCTCGACCGCACCGTTCTGGTCTATGGCCA
 GATGAACGAGCCCCCGGGGCCCGCTGGCGGGTGCCCATGACGGCGCTGACCATCGCCGAA
 TATTTCCGCGACGAGAAGCACCAGAACGTCCTGCTGCTGATGGACAACATCTTCCGCTTCGTC
 CAGGCGGGGGCGGAGGTCTCCGGCCTTTTGGGCCGTCCGCCCTCCCGGGTGGGATACCAGC
 CGACGCTGGCGAGCGAGGTGGCGGCGCTCCAGGAACGCATCACCTCCGTGGGCGAGGCCTC
 35 GGTGACCGCCATCGAGGCGGTCTACGTGCCGGCGGATGACTTCACCGATCCCGCCGTGACCA
 CCATCGCCGCCACGTGGATTCCATGGTGGTGCTCTCCCGCGCCATGGCGGCGGAGGGCAT
 GTATCCGGCGGTGGACCCCATCTCCTCCTCGTCGGTGCTGCTCGACCCGCTCATCGTGGGGG

ACGAGCATGCGCGCGTCGCCAACGAGGTGCGCCGGACCATCGAGCATTATCGCGAGCTTCAG
 GATGTGATCTCGCTGCTGGGCATGGAGGAATTGGGCACCGAGGATCGCCGCATCGTGGAGC
 GGGCGCGCCGGCTCCAGCGCTTCCTACCCAGCCCTTCACGGTCACCGAGGCCTTCACCGGC
 GTGCCCCGGCCGCTCGGTGGCCATCGCCGACACCATCGCCGGCTGCAGGATGATCCTGTCCG
 5 GCGCCTGCGACGACTGGCAGGAAAGCGCCCTCTACATGGTGGGCACCATCGACGAGGCCCCG
 CCAGAAGGAGGAGGCCGCTCGCGCCAAGGCGGGGCAGGGCGCCCCGGCCGGGACGGCAGC
 CGAGACGGCGGAGGCCGCCCGTGA

SEQ ID NO: 31:

10 Amino acid sequence of ATP synthase subunit beta atpD_1
 MAAADEEAQSAAGPASGRVVAVRGAVIDIAFAQPPLPLDDALLITDGRGGTVLVEVQSHMDRH
 TVRAIALQATTGLSRGLEAARVGGPVKVPVGDHVLGRLLDVTGAIGDKGGPLPADVPTRPIHHAP
 PSFAAQGGTSDLFRTGIKVIDLLAPLAQGGKAAMFGGAGVGKTVLVMELIHAMVASYKGISVFAG
 VGRSREGHEMLLDMTDSGVLDRTVLVYGQMNEPPGARWRVPMTALTIAEYFRDEKHQNVLLLM
 15 DNIFRFVQAGAEVSGLLGRPPSRVGYQPTLASEVAALQERITSVGEASVTAIEAVYVPADDFTDPA
 VTTIAAHVDSMVVLSRAMAAEGMPAVDPISSSSVLLDPLIVGDEHARVANEVRRRTIEHYRELQD
 VISLLGMEELGTEDRRIVERARRLRFLTQPFTVTEAFTGVPGRSVAIADTIAGCRMILSGACDDW
 QESALYMGVTIDEARQKEEAARAKAGQGAPAGTAAETAEAAP

20 SEQ ID NO: 32:

Nucleotide sequence of ATP synthase subunit beta atpD_2
 ATGGCGAACAAAGGTCGGACGCATCACCCAGATCATCGGCGCCGTCGTCGACGTGCAGTTCGA
 CGGGCATCTGCCGGCGATTCTCAACGCGATCGAGACCACCAACCAGGGCAACCGGCTGGTGC
 TCGAAGTGGCTCAGCATCTCGGCGAGAACACCGTGCCTGCATCGCCATGGATGCCACTGAA
 25 GGCCTGGTGCCTGGCCAGGAGGTGGCCGACACCGATGCGCCCATCCAGGTGCCCCTGGGCG
 CCGCCACCCTCGGCCGCATCATGAACGTGATCGGCGAGCCGGTGGACGAGCTGGGCCCCAT
 CGAGGGCGAAGCGCTGCGCGGCATCCATCAGCCGGCCCCCTCCTATGCGGAGCAGGCCACG
 GAAGCTGAGATCCTCGTCACCGGCATCAAGGTGGTGGATCTGCTGGCGCCCTATTCCAAGGG
 CGGCAAGGTGGGCCTGTTGCGCGGCGCCGGCGTGGGCAAGACCGTGCTCATCATGGAGCTG
 30 ATCAACAACGTGGCCAAGGCGCACGGCGGCTATTCCGTGTTGCGCGGCGTGGGTGAGCGCA
 CCCGCGAGGGCAACGACCTCTACCACGAGATGATCGAGTCCAACGTGAACAAGGACCCGCAC
 GAGAACAATGGCTCGGCGGCCGGTTCCAAGTGCGCCCTGGTCTATGGCCAGATGAACGAGCC
 GCCCGGCGCCCGCGCCCGCTGGCCCTCACCGGCCTCACCGTCGCCGAGCATTTCCGCGAC
 CAGGGCCAGGACGTGCTGTTCTTCGTGGACAACATCTTCCGCTTCACCCAGGCGGGCTCCGA
 35 GGTGTGCGGCGCTTCTCGGCCGCATCCCCTCGGCGGTGGGCTACCAGCCGACGCTGGCCACC
 GACATGGGCCAGCTGCAGGAGCGCATCACCACCACCAAGGGCTCCATCACCTCGGTGCA
 GGCCATCTACGTGCCGGCGGACGATCTGACCGATCCGGCGCCGGCCGCCTCCTTCGCCCATC

TGGACGCCACCACGGTGCTGTCGCGCTCCATCGCGGAGAAGGGCATCTACCCGGCGGTGGA
 TCCGCTGGACTCCACCTCGCGCATGCTGTCTCCCGCCATCCTCGGCGACGAGCACTACAACAC
 CGCGCGCCAGGTGCAGCAGACCCTGCAGCGCTACAAGGCGCTCCAGGACATCATCGCCATCC
 TGGGCATGGACGAACTCTCCGAAGAGGACAAGCTCACCGTGGCCCCGCGCCCGCAAGATCGA
 5 GCGCTTCTCTCCCAGCCCTTCCACGTGGCCGAGGTGTTACCGGTTCGCCCCGGCAAGCTGG
 TCGACCTCGCCGACACCATCAAGGGCTTCAAGGGCCTGGTGGACGGCAAGTACGACTACCTG
 CCCGAGCAGGCCTTCTACATGGTGGGCACCATCGAAGAAGCCATCGAGAAGGGCAAGAAGCT
 GGCGGCCGAGGCGGCCTGA

10 SEQ ID NO: 33:

Amino acid sequence of ATP synthase subunit beta atpD_2

MANKVGRITQIIGAVVDVQFDGHLPAILNAIETTQGNRLVLEVAQHLGENTVRCIAMDATEGLV
 RGQEVADTDAPIQVPVGAATLGRIMNVIGEPVDELGPIEGEALRGIHQPAPSYAEQATEAEILVTG
 IKVVDLLAPYSKGGKVGLFGGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYHEMI
 15 ESNVNKDPHENNGSAAGSKCALVYGQMNEPPGARARVALTGLTVAEHFRDQGQDVLFFVDNIFR
 FTQAGSEVSALLGRIPSAVGYQPTLATDMGQLQERITTTTKGSITSVQAIYVPADDLTDPAAPASF
 AHLDATTVLSRSIAEKGIYPAVDPLDSTSRMLSPAILGDEHYNTARQVQQTQLQRYKALQDIILG
 MDELSEEDKLTVARARKIERFLSQPFHVAEVFTGSPGKLVDLADTIKGFKGLVDGKYDYLPEQAFY
 MVGTIEEAIEKGKKLAAEAA

20

SEQ ID NO: 34:

Nucleotide sequence of ATP synthase gamma chain atpG_2

ATGGCGAGTCTGAAGGACCTGAGAAACCGCATTGCCTCGGTGAAGGCGACGCAGAAGATCAC
 CAAGGCGATGCAGATGGTCGCCGCGGCGAAGCTGCGTCGCGCCCAGGCGGCGGTGAAGC
 25 GGCCCGTCCCTATGCGGAACGCATGGAGACGGTGCTCGGAAATCTTGCCTCCGGCATGGTGG
 TGGGCGCGCAGGCGCCTGTTCTCATGACCGGGACGGGCAAGAGCGACACCCACCTGCTGCT
 GGTGTGCACCGGCGAGCGCGGCCTGTGCGGCGCCTTCAACTCGTCCATCGTGCGCTTCGCCC
 GCGAGCGGGCGCAGCTGCTGCTGGCCGAGGGCAAGAAGGTGAAAATCCTGTGCGTGGGCCG
 CAAGGGCCACGAGCAGCTGCGCCGCATCTACCCGGACAACATCATCGACGTGGTGGACCTGC
 30 GCGCGGTGCGCAACATCGGCTTCAAGGAGGCCGACGCCATCGCCCGCAAGGTGCTGGCCCT
 GCTCGATGAAGGCGCATTGACGTCTGCACGCTCTTCTACTCCACTTCAGGAGCGTGATCGC
 CCAGGTGCCGACGGCCAGCAGCTCATTCCGGCCACCTTCGACGAGCGGCCGGCCGTGCGC
 GATGCGCCGGTCTATGAATATGAGCCGGAGGAGGAGGAGATCCTCGCCGAGCTGCTGCCGC
 GCAACGTGGCGGTGCAGATCTTCAAGGCCCTCCTCGAGAACCAGGCTTCTTTCTATGGCTCCC
 35 AGATGAGCGCCATGGACAACGCCACGCGCAATGCGGGCGAGATGATCAAGAAGCAGACGCT
 CACCTACAACCGTACCCGCCAGGCCATGATCACGAAGGAACTCATCGAGATCATCTCCGGCG
 CCGAGGCCGTCTGA

SEQ ID NO: 35:

Amino acid sequence of ATP synthase gamma chain atpG_2

MASLKDLRNRIASVKATQKITKAMQMVA AAKLRRRAQAAAEARPYAERMETVLGNLASGMVVGA
 5 QAPVLMTGTGKSDTHLLLVCTGERGLCGAFNSSIVRFARERAQLLLAEGKKVKILCVGRKGHEQL
 RRIYPDNIIDVVDLRAVRNIGFKEADAIARKVLALLDEGAFDVCTLFYSHFRSVIAQVPTAQQLIPA
 TFDERPAVADAPVYEYEPEEEEEILAE LLPRNVAVQIFKALLENQASFYGSQMSAMDNATR NAGEM I
 KKQTLTYNRTRQAMITKELIEIISGA EAV

10 SEQ ID NO: 36:

Nucleotide sequence of ATP synthase subunit alpha atpA_2

ATGGACATTTCGAGCCGCTGAAATCTCTGCCATCCTGAAAGAGCAGATCCAGAATTTTCGGCCAG
 GAGGCGGAAGTCTCCGAGGTGGGTCAGGTTCTGTCCGTGGGTGACGGCATCGCGCGCGTCT
 ACGGCCTCGACAACGTCCAGGCGGGCGAGATGGTCGAGTTCGAGAACGGCACGCGCGGCAT
 15 GGCGCTGAACCTCGAGCTCGACAATGTCGGCATCGTGATCTTCGGTTCGACCGCGAGATCA
 AGGAAGGCCAGACCGTCAAGCGGACCGGCGCCATCGTGGACGCCCCCGTCGGCAAGGGCCT
 GCTCGGCCGCGTCGTGGACGCTCTCGGCAACCCGATCGACGGCAAGGGCCCGATCATGTTCA
 CCGAGCGTCGCCGGGTGACGTGAAGGCGCCGGGCATCATCCCGCGCAAGTCGGTGACAGA
 GCCCATGCAGACCGGCCTGAAGGCCATCGATGCGCTCATCCCCATCGGCCGCGGCCAGCGC
 20 GAGCTCATCATCGGCGACCGCCAGACCGGCAAGACCGCCGTGGCGCTCGACTCGATCCTGAA
 CCAGAAGCCCATCAACCAGGGCGACGACGAGAAGGCCAAGCTCTACTGCGTCTATGTCGCGG
 TGGGCCAGAAGCGTTCCACTGTCGCGCAGTTCTGTGAAGGTGCTCGAGGAGCACGGCGCGCT
 GGAATATTCCATCGTCGTCGCCGCCACCGCCTCGGACGCGGCCCCCATGCAGTTCCTGGCGC
 CGTTCACCGGCACCGCCATGGGCGAGTATTTCCGCGACAACGGCATGCACGCCCTCATCATC
 25 CATGATGACCTGTCCAAGCAGGCCGTGGCCTACCGCCAGATGTCGCTGCTGCTGCGCCGCCC
 GCCGGGCCGCGAGGCCTATCCCGGCGATGTGTTCTACCTGCACTCCCGCCTCTTGAGCGCG
 CCGCCAAGCTCAATGACGAGCACGGCGCCGGCTCGCTGACCGCCCTGCCGGTGATCGAGAC
 CCAGGCCAACGACGTGTCGGCCTACATCCCGACCAACGTGATCTCCATACCGACGGTCAGA
 TCTTCCTTGAATCCGATCTGTTCTACCAGGGCATCCGCCCCGGCGGTGAACGTGGGCCTGTG
 30 GTGTCGCGCGTGGGCTCTTCGGCCCAGATCAAGGCGATGAAGCAGGTGGCCGGCAAGATCA
 AGGGCGAGCTCGCCCAGTATCGCGAGCTGGCGGCCTTCGCCAGTTCGGTTCGGACCTGGA
 CGCGGCCACCCAGAAGCTGCTGAACCGCGGCGCCCGCCTCACCGAGCTGCTGAAGCAGAGC
 CAGTTCTCGCCCCTCAAGGTGGAGGAGCAGGTGGCGGTGATCTATGCCGGCACCAATGGCTA
 TCTCGATCCGCTGCCGGTCTCCAAGGTGCGCGAGTTCGAGCAGGGTCTGCTCCTGTGCTGCG
 35 GCTCGCAGCATCCGGAGATCCTGGACGCCATCCGCACGTCCAAGGAGCTTCCAAGGACACC
 GCCGAGAAGCTGACGAAGGCCATCGACGCCTTCGCCAAGAGCTTCTCCTGA

SEQ ID NO: 37:

Amino acid sequence of ATP synthase subunit alpha atpA_2

MDIRAAEISAILKEQIQNFGQAEVSEVGQVLSVGDGIARVYGLDNVQAGEMVEFENGTRGMAL
 NLELDNVGIVIFGSDREIKEGQTVKRTGAIVDAPVGKGLLGRVVDALGNPIDGKGPMFTERRRV
 5 DVKAPGIIPRKSVEHEPMQTGLKAIDALIPIGRGQRELIIGDRQTGKTAVALDSILNQKPINQGDDE
 KAKLYCVYVAVGQKRSTVAQFVKVLEEHLGALEYSIVVAATASDAAPMQFLAPFTGTAMGEYFRDN
 GMHALIIHDDLKQAVAYRQMSLLLRPPGREAYPGDVFYLSRLLERAALKLNDEHGAGSLTALP
 VIETQANDVSAYIPTNVISITDGQIFLESDFYQGIRPAVNVGLSVSRVGSSAQIKAMKQVAGKIK
 GELAQYRELAFAQFGSDLDAATQKLLNRGARLTELKQSQFSPLKVEEQVAVIYAGTNGYLDPLP
 10 VSKVREFEQGLLLSLRSQHPEILDAIRTSKELSKDTAEKLTAKIDAFKSF

SEQ ID NO: 38:

Nucleotide sequence of ATP synthase subunit delta atpH

GTGGCGGAAACGATCGTGTCAGGCATGGCGGGACGCTATGCGACCGCGCTGTTGAGCTGG
 15 CGGACGAAGCCGGTGCCATCGATTCCGTCCAGGCGGATCTTGATCGCCTGTCCGGCCTTCTG
 GCCGAGAGCGCGGATCTGGCGCGGCTGGTCAAGAGCCCGGTCTTCACCGCCGAGCAGCAGC
 TCGGCGCGATGGCGGCCATTCTCGATCAAGCAGGCATTTCCGGCCTTGCGGGCAAATTCGTG
 AAGCTGGTGGCGCAGAACCGCCGCCTGTTGCACTGCCGCGCATGATTGCCGAATACGCCGT
 CCTGGTGGCCCGGAAGAAGGGCGAGACCTCGGCGAGCGTGACCGTTGCCACCCCCCTGAGC
 20 GATGAGCATCTGGCCACGCTCAAGGCGGCCCTGGCTGAAAAGACCGGCAAGGACGTGAAGC
 TCGACGTCACCGTCGATCCGTCCATCCTCGGTGGTCTCATCGTGAAGCTCGGCTCGCGCATG
 GTCGATGCTTCCCTGAAGACCAAATCAATTCTATCCGGCATGCGATGAAAGAGGTCCGCTGA

SEQ ID NO: 39:

25 Amino acid sequence of ATP synthase subunit delta atpH

MAETIVSGMAGRYATALFELADEAGAIIDSVQADLDRSLGSLAESADLARLVKSPVFTAEQQLGAM
 AAILDQAGISGLAGKFVKLVQNRRLFALPRMIAEYAVLVARKKGETSASVTVATPLSDEHLATLK
 AALAEKTGKDVKLDVTVDPSILGGLIVKLGSRMVDASLTKLNSIRHAMKEVR

30 SEQ ID NO: 40:

Nucleotide sequence of ATP synthase subunit b atpF_2

ATGACCGAAATGGAACCTGGCTGAGCTCTGGGTCGCCATCGCCTTCCTGGTTTTCTAGGCCTC
 CTGATCTATGCGGGCGCCACCGCGCCATCGTCTCCGCCCTGGATTCCCGCGGCTCGCGCAT
 CGCCTCGGAACTGGAGGAGGCCCGTCGGCTCAAGGAAGAGGCCAGAGCTGGTGGCCGAA
 35 TTCAAGCGCAAGCAGCGCGAGGCCGAGGCCGAGGCCGAATCCATCGTCACCGGCGCCAAGG
 CCGAGGCCGAGCGCCTCGCCGCCGAGGCCAAGGCGAAGATCGAGGATTTCGTACCCGCCG
 CACCAAGATGGCCGAGGACAAGATCGCCAGGCCGAGCATCAGGCTCTGGCGGACGTGAAG

TCCATCGCCGCCGAGGCGGCGGCCAAGGCGGCCGAGGTGATCCTCGGCGCCCAGGCCACCG
GCGCGGTGGCGGAGCGTCTGCTGTCGGGCGCCATCTCCGAGGTCAAGACCAAGCTCAACTG
A

5 SEQ ID NO: 41:

Amino acid sequence of ATP synthase subunit b atpF_2

MTEMELAE LWVIAFLVFVGLLIYAGAHRAIVSALDSRGSRIASELEEARRLKEEAQKLVAEFKRK
QREAEAEASIVTGAKAEAERLAAEAKAKIEDFVTRRTKMAEDKIAQAEHQALADVKSIAAEAAA
KAAEVILGAQATGAVAERLLSGAISEVKTCLN

10

SEQ ID NO: 42:

Nucleotide sequence of ATP synthase subunit b' atpG_3

ATGATGATTGCATGGAAGCGGACCTTCGCAGTCGTGACCTTCGGGGCCGCCCTGATGGCCAT
GCCCCTCGCGGGCGTGGTCGCAGCTGAGACTTCTCCCGCTCCGGCGGCAGTGGCGCAGGCC
15 GATCATGCGGTGCCACCGAGGCGGCCGCGCCAGGGCACCGCCGATGCGGCCCATGCCGCCG
CGCCGGGCGAGGCCGCCCATGGTGGCGCGGCCAAGCACGAAACCCATTTCCCGCCCTTCGA
CGGCACCACCTTCGCCTCCCAGTTGCTGTGGCTCGCCGTCACCTTCGGCCTGCTTTACTACCT
CATGAGCAAGGTCACGCTGCCGCGCATCGGCCGCATCCTGGAAGAGCGCCACGACCGCATC
GCCGATGATCTGGAGGAAGCCTCCAAGCATCGCGCCGAGAGCGAGGCCGCCAGCGGGCCT
20 ATGAGAAGGCGCTGAGCGAGGCCCGCGCGAAGGCCCATTCATCGCCGCGGAAACCCGCGA
CCGCCTTGCCGCCACGCCGACACCAACCGCAAGGCGCTGGAGAGCGAGCTCACCGCCAAG
CTGCAGGCGGCCGAGGAGCGCATCGCCACCACCAAGAGCGAAGCCCTCACCCATGTGCGCG
GCATCGCGGTGGACGCCACCCAATCCATCGTCTCCACCCTCATCGGTGTCGCGCCCCGCGGCG
GCCGACGTGGAAAAAGCGGTGGACGGCGCCCTGTCCCAGCACGGCCAGGCCTGA

25

SEQ ID NO: 43:

Amino acid sequence of ATP synthase subunit b' atpG_3

MMIAWKRTFAVVTFGAALMAMPVAGVVAAETSPAPAAVAQADHAVPTEAAGQGTADAAHAAAP
GEAAHGGAACHETHFPFDGTTFASQLLWLAVTFGLLYLMSKVTLPRIGRILEERHDRIADDLEE
30 ASKHRAESEAAQRAYEKALSEARAKAHSIAAETRDRLAAHADTNRKALESELAKLQAAEERIATT
KSEALTHVRGIAVDATQSIVSTLIGVAPAAADVEKAVDGALSQHGQA

SEQ ID NO: 44:

Nucleotide sequence of ATP synthase subunit c atpE_2

35 ATGGAAGCGGAAGCTGGAAAGTTCATCGGTGCCGGCCTCGCCTGCCTCGGCATGGGTCTCGC
TGGCGTCGGCGTCGGTAACATCTTCGGTAACTTCCTCTCCGGCGCCCTGCGCAACCCGTCCG

CTGCCGACGGCCAGTTCGCCCCGCGCCTTCATCGGCGCCGCCCTCGCGGAAGGTCTCGGCATC
TTCTCGCTGGTCGTTGCGCTCGTCCTGCTGTTCTGCTGGCCTGA

SEQ ID NO: 45:

5 Amino acid sequence of ATP synthase subunit c atpE_2

MEAEAGKFIGAGLACLMGLAGVGVGNIFGNFLSGALRNPSAADGQFARAFIGAALAEGLGIFSL
VVALVLLFVA

SEQ ID NO: 46:

10 Nucleotide sequence of ATP synthase subunit a atpB_2

ATGACCGTCGATCCGATCCACCAGTTCGAGATCAAGCGCTACGTGGATCTGCTGAACGTCGG
CGGTGTCCAGTTCTCCTTCACCAACGCAACGGTGTTCATGATTGGCATCGTCCTGGTGATTTT
CTTCTTCCTGACTTTTCGCGACACGCGGTGCGACCCTTGTGCCGGGCCGGATGCAGTCGGCGG
CGGAGCTGAGCTACGAGTTCATCGCCAAGATGGTGCGCGACGCGGCCGGCAGCGAGGGAAT
15 GGTGTTCTTTCCCTTCGTCTTCTCGCTCTTCATGTTCTGTGCTGGTGGCGAACGTATTGGGGCT
CATCCCCTACACCTTCACGGTGACCGCCACCTCATCGTCACCGCCGCCCTGGCGGCGACGG
TGATCCTCACCGTCATCATCTACGGCTTCGTGCGGCACGGCACCCACTTCCTGCACCTGTTTCG
TGCCGTGCGGGCGTGCCGGGCTTCCTCCTGCCCTTCCTCGTGGTGATCGAGGTGGTGTCTGTT
CTGTCGCGGGCCATCAGCCTCTCGCTGCGTCTGTTCCGCAACATGCTGGCGGGCCACATCGC
20 CCTCAAGGTGTTGCCTTCTTCGTGCTGGGACTGGCCTCGGCCGGCGCGATCGGCTGGTTCG
GCGCCACCCTGCCCTTCTTCATGATCGTGGCGCTCACCGCGCTGGAGCTGCTGGTGGCGGTG
CTGCAGGCCTACGTGTTGCGGGTGTGACCTCGATCTACCTCAACGACGCCATCCATCCCGGC
CACTGA

25 SEQ ID NO: 47:

Amino acid sequence of ATP synthase subunit a atpB_2

MTVDPIHQFEIKRYVDLLNVGGVQFSFTNATVFMIGIVLVIFFFLTFA TRGRTLVPGRMQSAAELSY
EFI AKMVRDAAGSEGMVFFPFVFSLFMFVLVANVLGLIPYFTVT AHLIVTAALAATVILT VIIYGFV
RHGTHFLHLFVPSGVPGLLPFLV VIEVVSFLSRPISLSRLFANMLAGHIALKVFAFFV VGLASAGA
30 IGWFGATLPFFMIVALTALELLVAVLQAYVFAVLTSIYLND AIHPGH

SEQ ID NO: 48:

Nucleotide sequence of ATP synthase protein I atpI

ATGTCCGAGCCGAATGATCCATCCCGCAGGGACGGTGCGAAGGCGAAAGACGAGACGCAGG
35 ACTCCCGGCCCGGTGAGGCGGATCTTGCTCGGCGCCTCGATGCGCTCGGCACCTCCATCGGT
CAGGTCAAGTCCAGAAGCGGGGAGCCCGCGGCGACGCCGCGCAAGGACACCTCCTCGGCCT
CCGGCGCGGCCCTGGCGTTTCGGCTGGGCGCCGAGTTTGTTTCAGGCGTGCTGGTGGGCTC

GCTCATCGGCTACGGGTTGGATTATGCGTTTGCGATTTGCCCTGGGGGCTGATCGCCTTCAC
 GCTGATCGGCTTTGCCGCCGGCGTCCTGAACATGCTGCGCGTGCGGAACAGCGATGCCAAGC
 GCCACAGCGCGGACAGGTGA

5 SEQ ID NO: 49:

Amino acid sequence of ATP synthase protein I atpI

MSEPNDPSRRDGAKAKDETQDSRPGEADLARRLDALGTSIGQVKSRSGEPAATPRKDTSSASG
 AALAFRLGAEFVSGVLVGSLIGYGLDYAFAISPWGGLIAFTLIGFAAGVLNMLRVANSDAKRHSADR

10 In another embodiment, the bacterial strain used in the process of the invention
 comprises a gene encoding a nitrogenase molybdenum-iron protein alpha chain nifD_1
 having the sequence set forth in SEQ ID NO:51 or a sequence having more than 60%
 identity, e.g. more than 70% identity, such as more than 92% identity, such as more
 than 95% identity, e.g. more than 96% identity, such as more than 97% identity, e.g.
 15 more than 98%, such as more than 99% sequence identity to the sequence set forth in
 SEQ ID NO:51.

In another embodiment, the bacterial strain used in the process of the invention
 comprises a gene encoding nitrogenase molybdenum-iron protein alpha chain nifD_2
 20 having the sequence set forth in SEQ ID NO:53 or a sequence having more than 60%
 identity, e.g. more than 98%, such as more than 99% sequence identity to the sequence
 set forth in SEQ ID NO:53.

In another embodiment, the bacterial strain used in the process of the invention
 25 comprises a gene encoding a nitrogenase molybdenum-iron protein beta chain nifK_1
 having the sequence set forth in SEQ ID NO:55 or a sequence having more than 87%
 identity, e.g. more than 90% identity, such as more than 95% identity, e.g. more than
 96% identity, such as more than 97% identity, e.g. more than 98%, such as more than
 99% sequence identity to the sequence set forth in SEQ ID NO:55.

30

In another embodiment, the bacterial strain used in the process of the invention
 comprises a gene encoding a nitrogenase molybdenum-iron protein beta chain nifK_2
 having the sequence set forth in SEQ ID NO:57 or a sequence having more than 95%
 identity, e.g. more than 96% identity, such as more than 97% identity, e.g. more than
 35 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID
 NO:57.

In another embodiment, the bacterial strain used in the process of the invention comprises a gene encoding a nitrogenase iron protein nifH having the sequence set forth in SEQ ID NO:59 or a sequence having more than 98.5% sequence identity to the sequence set forth in SEQ ID NO:59.

5

SEQ ID NO: 50:

Nucleotide sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_1

ATGAGTTCGCTCTCCGCCACTATTCAACAGGTCTTCAACGAGCCGGGCTGCGCGAAGAACCA
 GAATAAGTCCGAGGCGGAGAAGAAGAAGGGCTGCACCAAGCAGCTGCAACCCGGCGGAGCG
 10 GCCGGCGGCTGCGCGTTTCGACGGCGCGAAGATCGCGCTCCAGCCCTTGACCGACGTCGCCC
 ACCTGGTGCACGGCCCCATCGCCTGCGAAGGCAATTCCTGGGACAATCGTGGCGCCAAGTCC
 TCCGGCTCGAACATCTGGCGCACC GGCTTACCACGGACATCAACGAAACCGACGTGGTGT
 CGGCGGCGAGAAGCGTCTGTTCAAGTCCATCAAGGAAATCATCGAGAAGTACGACCCGCCGG
 CCGTCTTCGTCTATCAGACCTGCGTCCCCGCCATGATCGGCGACGACATCGACGCGGTGTGC
 15 AAGGCGGCCAGGGAGAAGTTCGGAAGCCGGTGATCCCGATCAATCCCCCGGCTTCGTGG
 GGCCGAAGAATCTCGGCAACAAGCTCGCCGGCGAGGCGCTCCTCGACCATGTGATCGGCACC
 GAGGAGCCCGATTACACGACGGCCTACGACATCAACATCATCGGCGAATAACAATCTCTCCGG
 CGAGTTGTGGCAGGTGAAGCCGCTGCTGGACGAGCTGGGCATCCGCATCCTCGCCTGCATCT
 CCGGCGACGGGAAGTACAAGGATGTGGCGTCCTCCCACCGCGCCAAGGCGGCGATGATGGT
 20 GTGCTCCAAGGCCATGATCAACGTGGCCCGCAAGATGGAGGAGCGCTACGACATCCCCTTCT
 TCGAAGGCTCCTTCTACGGCATCGAGGATAGCTCCGATTCCCTGCGCGAGATTGCGCGCATG
 CTCATCGAGAAGGGCGCCGATCCGGAGCTGATGGACCGCACCGAGGCGCTGATTGAGCGGG
 AAGAGAAGAAGGCGTGGGACGCCATCGCCGCCTACAAGCCCCGCTTCAAGGACAAGAAGGT
 GCTGCTCATCACCGGCGGCGTGAAATCCTGGTCGGTGGTGGCAGCGCTCCAGGAAGCCGGC
 25 CTCGA ACTGGTGGGCACCTCGGTGAAGAAGTCCACCAAGGAGGACAAGGAGCGCATCAAGG
 AACTGATGGGCCAGGACGCCACATGATCGACGACATGACGCCCCGCGAAATGTACAAGATG
 CTGAAGGACGCCAAGGCGGACATCATGCTCTCGGGCGGGCGCTCGCAATTCATCGCGCTCAA
 GGCCGCCATGCCCTGGCTCGACATCAACCAGGAGCGCCACCACGCCTATATGGGCTATGTGG
 GCATGGTGAAGCTGGTCGAGGAGATCGACAAGGCGCTCTACAATCCCGTGTGGGAACAGGT
 30 GCGCAAGCCCGCCCCGTGGGAAAATCCGGAAGACACCTGGCAGGCCCGTGCGCTCGCCGAA
 ATGGAGGCGGAGGCCGCGCTCGCCGCCGATCCGGTGCGCGCGGAAGAGGTGCGCCGG
 TCCAAGAAGATCTGCAATTGCAAGAGCGTCGACCTCGGAACCATTGAGGACGCCATCAAGGC
 TCACGCGCTGACCACCGTGGAGGGTGTGCGAGAGCACACCAATGCCTCGGGAGGCTGCGGA
 GCCTGCAGCGGGCGGATCGAGGAGATCTTCGAGGCCGTGGGCGTTGTGCGCCGCCCGCCTC
 35 CCGCGGAGGCCGCCCGTCTCCGCAGGAGATCGCGCCCGATCCGCTCGCTGCGGAGGAAAA
 GCGCCGCGCCAAGAAGGCCTGCGGCTGCAAGGAGGTAGCGGTGCGCACCATGAGGATGCC
 ATCCGCGCCAAGGGTCTGCGAAACATCGCGGAGGTGCGTGCGGCCACCGATGCCAACACCG

GCTGCGGCAATTGCCAGGAGCGGGTGGAGGGCATCCTCGACCGGGTTCTCGCCGAGGCGGC
CTCAGAACTCCAGGCGGCGGAATAG

SEQ ID NO: 51:

5 Amino acid sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_1
MSSLSATIQQVFNEPGCAKNQNKSEAEKKKGCTKQLQPGGAAGGCAFDGAKIALQPLTDVAHLV
HGPIACEGNSWDNRGAKSSGSNIWRTGFTTDINETDVVFGGEKRLFKSIKEIIEKYDPPAVFVYQ
TCVPAMIGDDIDAVCKAAREKFGKPVIPINSPGFVGPKNLGNKLAGEALLDHVIGTEEPDYTTAYD
INIIGEYNLSGELWQVKPLLELGIRILACISGDGKYKDVASSHRAKAAMMVCSKAMINVARKME
10 ERYDIPFFEGSFYGIEDSSDSLREIARMLIEKGADPELMDRTEALIEREEKKAWDAIAAYKPRFKDK
KVLLITGGVKSWSVVAALQEAGLELVGTSVKKSTKEDKERIKELMGQDAHMIDDMTPREMYKML
KDAKADIMLSGGRSQFIALKAAMPWLDINQERHHAYMGYVGMVKLVEEIDKALYNPVWEQVRKP
APWENPEDTWQARALAEMEAEEAALAADPVRAEEVRRSKKICNCKSVDLGTIEDAIIKAHALTTVE
GVREHTNASGGCGACSGRIEEIFEAVGVVAAPPPAEAAPSPQEIAPDPLAAEEKRRRAKKACGCKE
15 VAVGTIEDAIRAKGLRNIAEVRAATDANTGCGNCQERVEGILDRVLAEEASELQAAE

SEQ ID NO: 52:

Nucleotide sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_2
ATGAGTGTGCGCACAGTCCCAGAGCGTCGCCGAGATCAAGGCGCGCAACAAGGAACTCATCGA
20 AGAGGTCCTCAAGGTCTATCCCGAGAAGACCGCCAAGCGCCGCGCCAAGCACCTGAACGTCC
ACGAAGCCGGCAAGTCCGACTGCGGCGTGAAGTCCAACATCAAGTCCATCCCGGGCGTGATG
ACCATCCGCGGTTGCGCTTATGCCGGCTCCAAGGGTGTGGTGTGGGGTCCCATCAAGGACAT
GATCCACATCTCCACGGCCCGGTGGGCTGCGGCCAGTATAGCTGGGCCGCCCCGCCGCAACT
ACTATATCGGCACGACCGGCATCGACACCTTCGTGACGATGCAGTTCACCTCCGACTTCCAGG
25 AGAAGGACATCGTCTTCGGCGGCGACAAGAAGCTCGCCAAGATCATGGACGAGATCCAGGAG
CTGTTCCCGCTGAACAACGGCATCACCGTTCAGTCCGAGTGCCCCATCGGCCTCATCGGCGA
CGACATCGAGGCCGTCTCCAAGCAGAAGTCCAAGGAGTATGAGGGCAAGACCATCGTGCCGG
TGCGCTGCGAGGGCTTCCGCGGCGTGTCCAGTCCCTGGGCCACCACATCGCCAACGACGCC
ATCCGCGATTGGGTGTTGACAAGATCGCGCCCGACGCCGAGCCGCGCTTTGAGCCGACCCC
30 GTACGACGTCGCCATCATCGGCGACTACAATATCGGTGGTGACGCCTGGTCGTCCCGTATCCT
CCTGGAGGAGATGGGCCTGCGCGTGATCGCCAGTGGTCCGGCGACGGTTCGCTCGCTGAG
CTGGAGGCCACCCCGAAGGCCAAGCTCAACGTGCTGCACTGCTACCGCTCCATGAACTACAT
CTCGCGCCACATGGAAGAGAAGTACGGTATCCCGTGGTGCGAGTACAATTCTTCGGTCCTTC
CAAGATCGCCGAGTCCCTGCGCAAGATCGCCAGCTACTTCGACGACAAGATCAAGGAAGGCG
35 CGGAGCGCGTCATCGCCAAGTATCAGCCGCTCATGGATGCGGTGATCGCGAAGTATCGTCCC
CGCCTCGAGGGCAAGACCGTGATGCTGTACGTGGGCGGCCTGCGTCCCCGTACGTCATCG
GCGCCTACGAGGACCTGGGCATGGAAGTGGTCGGCACGGGCTACGAGTTCGCCCATAAACGA

CGACTACCAGCGCACCGCCCAGCACTACGTCAAGGATGGCACCATCATCTATGACGACGTGA
 CCGGCTACGAGTTCGAGAAGTTCGTGCGAGAAGATCCAGCCGGACCTGGTCGGTTCGGGCATC
 AAGGAAAAGTACGTCTTCCAGAAGATGGGCGTGCCGTTCCGCCAGATGCACTCCTGGGACTA
 CTCGGGCCCCGTACCACGGCTATGACGGCTTCGCGATCTTCGCGCGCGACATGGACATGGCCA
 5 TCAACAGCCCCGTGTGGAAGATGACCCAGGCTCCGTGGAAGAGCGTCCCCAAGCCGACGATG
 CTCGCGGCTGAATGA

SEQ ID NO: 53:

Amino acid sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_2
 10 MSVAQSQSVAEIKARNKELIEEV LKVYPEKTAKRRAKHLNVHEAGKSDCGVKSNIKSIPGVMTIR
 GCAYAGSKGVVWGPIKDMIHISHGPGVCGQYSWAARRNYIGTTGIDTFVTMQFTSDFQEKDIV
 FGDKKLAKIMDEIQELFPLNNGITVQSECPIGLIGDDIEAVSKQKSKEYEGKTIVPVRCEGFRGV
 SQSLGHHIANDAIRDWVFDKIAPDAEPRFEPTPYDVAIIGDYNIGGDAWSSRILLEEMGLRVIAQ
 WSGDGS LAELEATPKAKLNLH CYRSMNYISRHMEEKYGIPWCEYNFFGPSKIAESLRKIASYFD
 15 DKIKEGAERVI AKYQPLMDAVIAKYRPRLEGKTVMLYVGGLRPRHVIGAYEDLGMEVVGTGYEFA
 HNDDYQRTAQHYVKDGTIIYDDVTGYEFKFEKIQPDLVGSIGIKEKYVFQKMGVPFRQMHSWD
 YSGPYHGYDGF AIFARDMDMAINSPVWKMTQAPWKSVPKPTMLAAE

SEQ ID NO: 54:

Nucleotide sequence of Nitrogenase molybdenum-iron protein beta chain nifK_1
 20 ATGGCCACCGTTTCCGTCTCCAAGAAGGCCTGCGCGGTCAACCCCTCAAGATGAGCCAGCC
 GGTGGGCGGCGCGCTCGCCTTCATGGGCGTGCGCAAGGCCATGCCGCTGCTGCACGGCTCG
 CAGGGCTGCACCTCCTTCGGCCTGGTGCTGTTCTGCGCCACTTCAAGGAAGCCATCCCCAT
 GCAGACCACCGCCATGAGCGAGGTGGCGACGGTCTGGGCGGCCTTGAGAATGTGGAGCAG
 25 GCCATTCTCAACATCTACAATCGCACCAAGCCGGAGATCATCGGCATCTGCTCCACCGGCGTC
 ACCGAGACCAAGGGCGATGATGTCGACGGCTACATCAAGCTGATCCGGGACAAGTATCCCCA
 GCTGGCCGACTTCCCGCTGGTCTATGTCTCCACCCCGATTTCAAGGACGCCTTCCAGGACG
 GTTGGGAGAAGACCGTGGCGAAGATGGTGGAGGCGCTGGTGAAGCCCGCCGCCGACAAGCA
 GAAGGACAAGACCCGCGTCAACGTCCTGCCCGGCTGCCACCTCACGCCC GGCGATCTGGATG
 30 AGATGCGGACCATCTTCGAGGATTTGGGCTCACACCCTATTTCTGCGGATCTGGCCGGCT
 CGCTGGATGGGCATATCCCCGAGGACTTCTCGCCCACCACCATCGGCGGCATCGGCATCGAT
 GAGATCGCCACCATGGGCGAGGCGGCCACACCATCTGCATCGGCGCGCAGATGCGCCGGG
 CGGGCGAGGCCATGGAGAAGAAGACCGGCATTCCCTTCAAGCTGTTTCGAGCGCCTGTGCGG
 CCTGGAGGCGAACGACGCCTTCATCATGCACCTGTGCGAGATCTCCGGCCGGCCGGTGCCGG
 35 TGAAGTATCGCCGGCAGCGGGGCCAGCTGGTGGATGCCATGCTGGACGGCCACTTCCATCTG
 GGCGGTGCGCAAGGTGGCCATGGGGGCGGAGCCGGACCTGCTCTACGACGTGGGCTCCTTCC
 TGCACGAGATGGGCGCCCACATCCTTTCCGCGGTACCAACCACCCAGTCGCCGGTGCTGGCG

CGCCTGCCTGCCGAGGAGGTGCTTATCGGCGACCTGGAGGATCTGGAGACCCAGGCGAAGG
 CGCGCGGATGCGATCTCCTGCTACCCATTCCCATGGGCGCCAGGCGGCGGAGCGCCTCCAC
 ATCCCCTTCTACCGGATCGGCATTCCCATGTTTGACCGGCTGGGGGCGGGGCATCTGTTGTC
 GGTGGGCTATCGCGGCACCCGCGACCTCATCTTCCATCTCGCCAACCTTGTGATCGCCGACCA
 5 CGAGGAAAATCACGAGCCGACGCCCCGACACCTGGGCCACCGGCCATGGCGAGCATGCCGCC
 GCCCCCACTTCCCATTGA

SEQ ID NO: 55:

Amino acid sequence of Nitrogenase molybdenum-iron protein beta chain nifK_1
 10 MATVSVSKKACAVNPLKMSQPVGGALAFMGVRKAMPLHGSQGCTSFGLVLFVRHFKEAIPMQT
 TAMSEVATVLGGLENVEQAILNIYNRTKPEIIGICSTGVTETKGDDVDGYIKLIRDKYPQLADFPLV
 YVSTPDFKDAFQDGWEKTVAKMVEALVKPAADKQKDKTRVNVLPGCHLTPGDLDEMRTIFEDFG
 LTPYFLPDLAGSLDGHIPEDFSPTTIGGIGIDEIATMGEEAHTICIGAQMRRAGEAMEKKTGIPFKL
 FERLCGLEANDAFIMHLSQISGRPVVKYRRQRGQLVDAMLDGHFHLGGRKVAMGAEPDLLYDV
 15 GSFLHEMGAHILSAVTTTQSPVLARLPAEEVLIGDLEDLETQAKARGCDLLLTHSHGRQAAERLHI
 PFYRIGIPMFDRLGAGHLLSVGYRGTRDLIFHLANLVIADHEENHEPTPDWTATGHGEHAAPTS
 H

SEQ ID NO: 56:

Nucleotide sequence of Nitrogenase molybdenum-iron protein beta chain nifK_2
 20 ATGCCACAAAATGCTGACAATGTGCTCGATCACTTCGAGCTCTTCCGTGGTCCCGAATACCAG
 CAGATGCTGGCCAATAAGAAAAAGATGTTTCGAGAACCCCCGCGATCCGGCCGAAGTCGAGCG
 CGTGCGGGAATGGGCGAAGACTCCTGAATACAAGGAGCTGAACTTCGCCCCGCGAGGCGCTC
 ACCGTGAATCCGGCCAAGGCTTGTACGCCGCTGGGCGCGGTGTTTCGTCGCCGTTCGGCTTCGA
 25 GAGCACGATCCCCTTCGTGCACGGCTCGCAGGGTTGCGTCGCGTATTACCGCTCGCACCTCT
 CCCGCCACTTCAAGGAGCCGTCCTCCTGCGTCTCCTCGTCCATGACCGAGGATGCGGCGGTG
 TTCGGCGGCCTCAACAACATGATTGACGGCCTCGCCAACACCTACAACATGTACAAGCCGAAG
 ATGATCGCCGTCTCCACCACCTGCATGGCGGAAGTCATCGGCGACGATCTGAACGCCTTCATC
 AAGACCGCGAAGGAAAAGGGCTCGGTTCCGGCCGAATACGACGTGCCCTTCGCCCACACCCC
 30 GCGTTCGTCGGCAGCCATGTACCGGCTACGACAATGCGCTCAAGGGCATCCTCGAGCACT
 TCTGGGACGGCAAGGCCGGCACCGCGCCGAAGCTGGAGCGCGTTCCTCAACGAGAAGATCAA
 CTTTCATCGGCGGCTTCGACGGCTACACCGTCGGCAACACTCGCGAAGTGAAGCGCATCTTCG
 AGGCGTTTCGGCGCCGATTACACCATCCTCGCCGACAATTCCGAAGTGTTTCGACACCCCGACC
 GACGGCGAGTTCCGCATGTATGACGGCGGCACGACCCTGGAGGACGCGGCGAACGCGGTGC
 35 ACGCCAAGGCCACCATCTCCATGCAGGAATACTGCACGGAGAAGACCCTGCCCATGATCGCC
 GGTTCATGGCCAGGACGTGGTCGCCCTCAACCACCCCGTGGGCGTGGGCGGCACCGACAAGT
 TCCTCATGGAGATCGCCCGCCTACCGGCAAGGAGATCCCCGAGGAGCTGACCCGCGAGCG

CGGCCGTCTCGTGGACGCTATCGCGGACTCTTCCGCGCACATCCACGGCAAGAAGTTCGCCA
TCTACGGCGATCCGGATCTGTGCCTGGGCCTCGCCGCGTTCCTGCTGGAGCTGGGCGCCGAG
CCGACCCATGTGCTGGCCACCAACGGCACCAAGAAGTGGGCCGAGAAGGTTTCAGGAACTGTT
CGACTCTTCGCCGTTTCGGCGCCAACTGCAAGGTCTATCCCGGCAAGGACCTGTGGCACATGC
5 GCTCGCTCCTGTTCGTGGAGCCGGTGGATTTTCATCATCGGCAACACCTACGGCAAGTATCTCG
AGCGCGACACGGGCACCCCGCTGATCCGTATCGGCTTCCCGGTGTTTCGACCGTCACCACCAC
CACCGCCGTCCGGTGTGGGGCTATCAGGGCGGCATGAACGTCCTGATCACGATCCTCGACAA
GATCTTTGACGAGATCGACCGCAACACCAACGTGCCGGCCAAGACCGACTACTCGTTCGACAT
CATTCGTTGA

10

SEQ ID NO: 57:

Amino acid sequence of Nitrogenase molybdenum-iron protein beta chain nifK_2

MPQNADNVLDHFELFRGPEYQQMLANKKKMFENPRDPAEVERVREWAKTPEYKELNFAREALTV
NPAKACQPLGAVFVAVGFESTIPFVHGSQGCVAYYRSHLSRHFKEPSSCVSSSMTEDAAVFGGLN
15 NMIDGLANTYNMYKPKMIAVSTTCMAEVIGDDLNAFIKTAKEKGSVPAEYDVPFAHTPAFVGSHV
TGYDNALKGILEHFWDGKAGTAPKLERVPNEKINFIGGFDGYTVGNTREVKRIFEAFGADYILAD
NSEVFDTPDGEFRMYDGGTTLEDAANAVHAKATISMQEYCTEKLPMIAGHGQDVVALNHPVG
VGGTDKFLMEIARLTGKEIPEELTRERGRILDADSSAHIHGKKFAIYGDPLCLGLAAFLLELGA
EPHVLATNGTKKWAQELFDSSPFGANCKVYPGKDLWHMRSLLFVEPVDFIIGNTYGKYLER
20 DTGTPLIRIGFPVFDHRHHHRPVGWYQGGMNVLITLIDKIFDEIDRNTNVPKTDYSFDIIR

SEQ ID NO: 58:

Nucleotide sequence of Nitrogenase iron protein nifH

GTGGAGTCCGGTGGTCCTGAGCCGGGCGTGGGCTGCGCCGGCCGCGGCGTGATCACCTCCA
25 TCAACTTCCTGGAGGAGAACGGCGCCTACGAGGACATCGACTATGTGTCCTACGACGTGCTG
GGCGACGTGGTGTGCGGCGGCTTCGCCATGCCATCCGCGAGAACAAAGGCGCAGGAAATCT
ACATCGTGATGTCCGGCGAGATGATGGCCATGTATGCGGCCAACAAACATCTCCAAGGGCATC
CTGAAGTATGCCAATTCCGGCGGCGTGCGCCTGGGCGGGCTGGTCTGCAACGAGCGCCAGA
CCGACAAGGAGCTGGAGCTGGCGGAGGCTCTGGCGAAGAAGCTCGGCACCGAGCTGATCTA
30 CTTCGTGCCGCGCGACAACATCGTGACGATGCCGAGCTGCGCCGCATGACAGTGATCGAGT
ATGCGCCCGATTCCGCCAGGCCAGCACTACCGGAACCTGGCCGAGAAGGTGCACGCCAAC
AAGGGCAACGGCATCATCCCGACCCGATCACCATGGACGAGCTGGAAGACATGCTCATGGA
GCACGGCATCATGAAGGCCGTGGACGAGAGCCAGATCGGCAAGACCGCCGCGGAGCTCGCC
GTCTGA

35

SEQ ID NO: 59:

Amino acid sequence of Nitrogenase iron protein nifH

MESGGPEPGVGCAGRGVITSINFLEENGAYEDIDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVM
 SGEMMAMYAANNISKGILKYANSSGGVRLGGLVCNERQTDKELELAEALAKKLGTELIYFVPRDNI
 VQHAELRRMTVIEYAPDSAQAQHYRNLAEKVHANKGNGIIPITMDELEDMLMEHGIMKAVDES
 QIGKTAAELAV

5

Downstream processing

In one embodiment, the process of the invention comprises the further step of harvesting biomass produced during the culture. Biomass can e.g. be harvested by sedimentation (settling based on gravity), filtration, centrifugation or flocculation.
 10 Flocculation may require the addition of a flocculation agent. Centrifugation may e.g. be carried out using a continuous flow centrifuge.

In one embodiment, the harvested biomass is subsequently dried. Drying can e.g. be performed using well known methods, including centrifugation, drum drying,
 15 evaporation, freeze drying, heating, spray drying, vacuum drying and/or vacuum filtration. The dried biomass may subsequently be used in a product, e.g. a food or feed product or feed or food ingredient.

In another embodiment, the cells of the harvested biomass are lysed. The lysate may
 20 in some embodiments be separated into insoluble and soluble fractions, either or both of which may subsequently be concentrated or dried, and subsequently be used in a product, e.g. a food or a feed product.

In one embodiment, biomass is harvested and proteins are isolated from said biomass,
 25 resulting in a protein fraction and a fraction comprising non-protein components. Depending on the method of protein isolation, the resulting fractions may be more pure or less pure. Thus, the term "protein fraction" means a fraction enriched in proteins. The protein fraction may still comprise significant amounts of other components and also significant amounts of protein may end up in the "fraction comprising non-protein
 30 components".

Isolation of proteins may be performed using any suitable method. For example, in one embodiment, proteins are isolated by breaking cells mechanically and separating protein from cell debris through one or more filtration steps, e.g. successive filtration through
 35 multiple filters with decreasing pore size. Mechanical breaking may be carried out using any suitable method, e.g. ball milling, sonication, homogenization, high pressure homogenization, mechanical shearing, etc. The resulting filtered protein fraction will be

enriched in proteins, but also still contain other smaller components. Protein may optionally be further purified from this fraction using any suitable method.

In another embodiment, a protein fraction is isolated by performing ethanol extraction followed by one or more filtration steps. Such methods are e.g. known from the preparation of soy bean proteins (see e.g. Chapter 5 "Soybean Protein Concentrates" in "Technology of production of edible flours and protein products from soybeans" by Berk
5
FAO Agricultural Services Bulletin No. 97 (1992)). The resulting protein fraction will be enriched in proteins, but also still contain other components. Protein may optionally be
10 further purified from this fraction using any suitable method.

In one embodiment, the process of the invention comprises the further step of hydrolysing the protein fraction obtained from the process of the invention to obtain amino acids and small peptides.

15

In one embodiment of the process of the invention, the process comprises the further step of producing a food or feed product from said biomass, from said protein fraction or from said fraction comprising non-protein components. Said further step may simply comprise incorporating said biomass, protein fraction or fraction comprising non-protein
20 components in a food or feed product, by adding it during the production of the food or feed product. In other embodiments, further purification or modification of the biomass or fraction thereof is performed during the course of its incorporation into a food or feed product.

25

The invention is further illustrated with the following, non-limiting, examples:

EXAMPLES

Example 1. Isolation of bacterial strain capable of chemoautotrophic growth

30 A sample of 50 mL containing soil and seawater was collected in a sterile falcon tube from the seashore of the Baltic sea in Naantali in Finland. Part of soil sample was mixed with 10 mL of mineral medium in a sterile Erlenmeyer flask. The medium consisted of 1 g/L NH_4OH , 0.23 g/L KH_2PO_4 , 0.29 g/L $\text{Na}_2\text{HPO}_4 \cdot 2 \text{H}_2\text{O}$, 0.005 g/L $\text{NaVO}_3 \cdot \text{H}_2\text{O}$, 0.2 g/L $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.5 g/L $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.01 g/L CaSO_4 , 0.00015 g/L $\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$,
35 0.005 g/L MnSO_4 , 0.0005 g/L $\text{ZnSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.0015 g/L H_3BO_3 , 0.001 g/L CoSO_4 , 0.00005 g/L CuSO_4 and 0.0001 g/L NiSO_4 prepared in tap water. The suspension of soil and medium was incubated in a shaking incubator in +30 °C temperature in a sealed

steel box that was flushed continuously with a gas mixture: 150 mL/min of N₂, 18 mL/min of H₂, 3 mL/min of O₂ and 6 mL/min of CO₂. The cultivation was refreshed in seven-day intervals by taking 1 mL of suspension, which was added in sterile conditions to 9 mL of medium in Erlenmeyer flask, and then placed back into the incubation box.

5 After the fourth dilution, there was no noticeable soil left in the suspension. The volume of the cell suspension was increased to 100 mL in order to grow biomass for bioreactor cultivation. The optical density (OD₆₀₀) of the suspension was 1.53 when it was inoculated to 190 mL of mineral medium in 15-vessel 200-mL parallel bioreactor system (Medicel Explorer, Medicel Oy, Finland). The cultivation conditions were 800 rpm

10 agitation, +30°C temperature and the pH was set to 6.8, controlling it with 1 M NaOH. Gas was fed through a sparger with a gas mixture consisting of 14 mL/min H₂, 3 mL/min O₂ and 6 mL/min CO₂. The head space of the reactor was flushed with 300 mL/min air. Continuous cultivation was fed with mineral medium 6 mL/h and cell suspension was drawn from the reactor via capillary keeping the volume constant at 200 mL. Cell

15 suspension drawn from the reactor was stored at +4°C. A sample was taken from the bioreactor automatically every day, and absorbance at 600 nm was measured to monitor the growth. After 498 hours of bioreactor cultivation, samples were drawn aseptically and suspension was diluted and plated to agar mineral medium plates containing the above minerals and 2% bacteriological agar. Plates were incubated in same conditions

20 as described above for the Erlenmeyer flasks. Colonies were then picked from agar plates and streaked to new agar plates in order to isolate one organism in one colony. This was repeated twice. Single colonies were picked and suspended into 200 µL of medium in a 96-well microtiter plate. The suspension was incubated at +30°C temperature and shaken 625 rpm in an EnzyScreen gas tight box that was flushed

25 continuously with 150 mL/min of N₂, 18 mL/min of H₂, 3 mL/min of O₂ and 6 mL/min of CO₂. The suspension from one well was transferred to an Erlenmeyer flask and supplemented with fresh medium. Volume was increased until there was enough biomass to perform a bioreactor cultivation. The organism was deposited in the VTT culture collection as VTT-E-193585.

30

16S rRNA sequencing of a sample demonstrated that the sample contained only one organism. The same sample was used for Illumina NextSeq sequencing providing 1x150 bp metagenomic shotgun sequences. Using Unicycler (Wick et al, 2017 PLoS computational biology 13:e1005595), the de novo assembly was made for metagenomic

35 sequences consisting of 101 contigs. The total genome length was 4,846,739 bp and the GC content was 67.9 %. Gene predictions and functional annotations were performed using Prokka (Seemann, 2014 Bioinformatics 30:2068). The genome

annotation produced 4,429 genes. Roary pan genomic alignment (Page et al, 2015 Bioinformatics 31:3691) grouped VTT-E-193585 among *Xanthobacter* species. The strain was therefore identified as a *Xanthobacter* sp., the closest genome being *Xanthobacter tagetidis*. Alignment-based calculation of average nucleotide identity that takes into account only orthologous fragments (OrthoANI) (Lee et al, 2016 Int J Syst Evol Microbiol 66:1100) gave the best match of 80.4% to *Xanthobacter tagetidis* (ATCC 700314; GCF_003667445.1), whereas the proposed species boundary cut-off is 95-96% (see e.g., Chun et al., 2018 Int J Syst Evol Microbiol, 68: 461-466). *Xanthobacter autotrophicus* Py2 gave a match of 79.6%, while the match for *Xanthobacter* sp. 91 was 79.0%. It could thus be concluded that the isolated bacterial strain deposited as VTT-E-193585 belongs to the Phylum: Proteobacteria; to the Class: Alpha Proteobacteria; and to the Order: Rhizobiales. The most probable Family is Xanthobacteraceae, and the Genus *Xanthobacter*. The VTT-E-193585 bacterial strain could not be assigned unequivocally to any known species.

15

A search for putative antimicrobial resistance genes was performed. The ABRicate (<https://github.com/tseemann/abrigate>) tool was used to search the genome against the Arg-Annot, NCBI, ResFinder, the ecOH, Megares and VFDB databases using blastn or blastp. A threshold of 50 % was set for both identity and coverage, both on nucleotide and protein level. Only two putative antimicrobial resistance genes were identified. These two genes did not contain amino-acid changes linked to antibiotics resistance and thus a resistant phenotype is not expected.

20

Example 2. Pilot cultivation and analysis of isolated bacterial strain

25

The isolated bacterial strain deposited as VTT-E-193585 was cultivated in a conventional 200-liter stirred tank bioreactor (MPF-U, Marubishi Ltd, Japan). Mixing was performed with Rushton-type impellers rotating at 400 rpm. Temperature in the cultivation was maintained at +30°C. pH was maintained at 6.8±0.2 by adding 8 M NaOH or 3.6 M H₃PO₄ by software control. Cultivation medium contained 1 g/L NH₄OH, 0.23 g/L KH₂PO₄, 0.29 g/L Na₂HPO₄ · 2 H₂O, 0.005 g/L NaVO₃ · H₂O, 0.2 g/L FeSO₄ · 7 H₂O, 0.5 g/L MgSO₄ · 7 H₂O, 0.01 g/L CaSO₄, 0.00015 g/L Na₂MoO₄ · 2 H₂O, 0.005 g/L MnSO₄, 0.0005 g/L ZnSO₄ · 7 H₂O, 0.0015 g/L H₃BO₃, 0.001 g/L CoSO₄, 0.00005 g/L CuSO₄ and 0.0001 g/L NiSO₄ prepared in tap water. A mixture containing 1.8-10.5 L/min hydrogen gas, 0.6-2.5 L/min oxygen gas and 1.8-5 L/min carbon dioxide gas was supplied constantly as the main source of energy and carbon. Dissolved oxygen level was maintained at 7.2±0.5 % by adjusting the gas mixture composition. The inoculum for the cultivation was prepared as described in Example 1. Growth was monitored by taking samples

35

manually and analysing the cell density as optical density by measuring absorbance at 600 nm (Ultrospec 2100 pro UV/visible spectrophotometer, Biochrom Ltd., England) and by measuring cell dry weight (CDW) by drying in oven overnight at 105°C. Optical density was also monitored by using an in situ absorbance probe (Trucell 2, Finesse Ltd, USA). A growth curve of the cultivation is presented in Figure 1. The maximum growth rate in batch phase was 0.06 h⁻¹. The maximum cell density was 4.5 g_{CDW}/L at 92 h. After 92 h of cultivation, feed of fresh cultivation medium as described above was started at a dilution rate of 0.01 h⁻¹. During the continued feed, the cell density was on average 2.9 g_{CDW}/L. Cultivation liquid was constantly collected to a cooled (+10°C) tank from which it was fed in 300-liter batches to a continuous centrifugal separator (BTPX-205, Alfa-Laval AB, Sweden). The cell-containing slurry collected from the separator was fed into an atmospheric double drum dryer (Buflovak 6x8 ADDD, Hebel process solutions LLC., USA), heated with 4 bar steam and drums rotating at 3.5 rpm. This resulted in dried cell powder with approximately 96% dry matter content. Analysis results of the dried cell powder are presented in Table 1 for the proximate composition, in Table 2 for the amino acid composition, in Table 3 for the fatty acid composition, and in Table 4 for the vitamin content. Analyses demonstrate that the dried cell powder has high protein content with all the essential amino acids. It also contains more unsaturated than saturated fatty acids and a lot of B-group vitamins. Peptidoglycan content was only 0.002 mg/g_{CDW} and lipopolysaccharide content was 0.01 mg/g_{CDW}. It would be beneficial that these concentrations would be as small as possible. In comparison, in a commercial lactic acid bacteria preparation analysed at the same time, the peptidoglycan content was 0.244 mg/g_{DW} and the lipopolysaccharide content was 0.015 mg/g_{DW}. Cytotoxicity and genotoxicity assays were performed using the supernatant samples of cultivation. No cytotoxicity against HepG2 or HeLa229 human cell lines was observed. No genotoxicity against *Escherichia coli* WP2 *trp*⁻ or CM871 *uvrA* *recA* *lexA* strains was observed.

Table 1. Analysis results of dried cell powder of isolated bacterial strain deposited as VTT-E-193585.

Parameter	Method	Unit	Value
Moisture	Drying at 103 °C	g/100g	3.3
Protein	Kjeldahl (N x 6,25)	g/100g	72.2
Fat	Weibull-Stoldt	g/100g	6.0
Saturated fatty acids	Calculation based on Table 3	g/100g	1.8
Mono-unsaturated fatty acids	Calculation based on Table 3	g/100g	3.8

Polyunsaturated fatty acids	Calculation based on Table 3	g/100g	0.4
Omega 3 fatty acids	Calculation based on Table 3	g/100g	< 0.01
Omega 6 fatty acids	Calculation based on Table 3	g/100g	0.4
Dietary fibres	Gravimetric	g/100g	10.5
Ash	Ashing at 550 °C	g/100g	5.8
Glucose	HPLC-ELSD	g/100g	< 0.15
Fructose	HPLC-ELSD	g/100g	< 0.1
Sucrose	HPLC-ELSD	g/100g	< 0.1
Lactose	HPLC-ELSD	g/100g	< 0.25
Maltose	HPLC-ELSD	g/100g	< 0.2
Total sugars	Calculation	g/100g	< 0.8
Carbohydrates	Calculation	g/100g	2.2
Energy	Calculation	kJ/100g	1572
Energy	Calculation	kcal/100g	373

Table 2. Amino acid composition of dried cell powder of isolated bacterial strain deposited as VTT-E-193585.

Parameter	Method	Unit	Value
Lysine	Ion chromatography	%	3.95
Methionine	Ion chromatography	%	1.60
Cystine	Ion chromatography	%	0.39
Aspartic	Ion chromatography	%	6.82
Threonine	Ion chromatography	%	3.47
Serine	Ion chromatography	%	2.75
Glutamic	Ion chromatography	%	8.84
Proline	Ion chromatography	%	3.14
Glycine	Ion chromatography	%	4.40
Alanine	Ion chromatography	%	6.94
Valine	Ion chromatography	%	4.96
Isoleucine	Ion chromatography	%	3.34
Leucine	Ion chromatography	%	6.08
Tyrosine	Ion chromatography	%	2.99
Phenylalanine	Ion chromatography	%	4.58
Histidine	Ion chromatography	%	1.66
Arginine	Ion chromatography	%	4.96

Tryptophan	HPLC	%	1.34
------------	------	---	------

Table 3. Fatty acid composition of dried cell powder of isolated bacterial strain deposited as VTT-E-193585.

Parameter	Method	Unit	Value
C16:0 (Palmitic acid)	GC-MS	%	24.8
C16:1 (Palmitoleic acid)	GC-MS	%	3.0
C18:0 (Stearic acid)	GC-MS	%	4.4
C18:1n9 (Oleic acid)	GC-MS	%	59.9
C18:2n6 (Linoleic acid)	GC-MS	%	6.1
C18:3n3 (alpha-Linolenic acid)	GC-MS	%	0.4

5 **Table 4. Vitamin content of dried cell powder of isolated bacterial strain deposited as VTT-E-193585.**

Parameter	Method	Unit	Value
Vitamin A (RE)	HPLC	IU/100 g	<100
Vitamin E (TE)	HPLC	mg/100 g	0.33
Vitamin D3	HPLC	IU/100 g	<10
Vitamin D2	HPLC	IU/100 g	21.6
Vitamin C	HPLC	mg/100 g	<1
Thiamine chloride Hydrochloride	LC-MS/MS	mg/100 g	0.9
Vitamin B1 (Thiamine)	Calculation	mg/100 g	0.708
Vitamin B2 (Riboflavin)	HPLC	mg/100 g	6.27
Pyridoxine hydrochloride	HPLC	mg/100 g	3.39
Vitamin B6 (Pyridoxine)	Calculation	mg/100 g	2.79
Vitamin B12	LC-MS/MS	µg/100 g	224
Choline chloride	LC-MS/MS	mg/100 g	14.3
Biotin	LC-MS/MS	µg/100 g	15.6
Folic acid	Microbiological	µg/100 g	1270
Niacin (Vitamin B3)	Microbiological	mg/100 g	23.2
Pantothenic acid	Microbiological	mg/100 g	6.53

Example 3. Cultivation of isolated bacterial strain on different nitrogen sources.

The isolated bacterial strain deposited as VTT-E-193585 was cultivated in a 15-vessel parallel bioreactor system at 200 mL volume (Medicel Explorer, Medicel Oy, Finland). Mixing was performed with Rushton-type impellers rotating at 800 rpm. The temperature in the cultivation was maintained at +30°C. pH was maintained at 6.8 by adding 1 M NaOH. The cultivation medium contained 0.23 g/L KH_2PO_4 , 0.29 g/L $\text{Na}_2\text{HPO}_4 \cdot 2 \text{H}_2\text{O}$, 0.005 g/L $\text{NaVO}_3 \cdot \text{H}_2\text{O}$, 0.2 g/L $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.5 g/L $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.01 g/L CaSO_4 , 0.00015 g/L $\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$, 0.005 g/L MnSO_4 , 0.0005 g/L $\text{ZnSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.0015 g/L H_3BO_3 , 0.001 g/L CoSO_4 , 0.00005 g/L CuSO_4 and 0.0001 g/L NiSO_4 prepared in tap water. Furthermore, the nitrogen source was varied in the cultivations so that four cultivations contained 18.7 mM NH_4OH , four cultivations contained 9.34 mM urea ($\text{OC}(\text{NH}_2)_2$), four cultivations contained 18.7 mM potassium nitrate (KNO_3), and three cultivations were left without nitrogen source in the medium. A mixture containing 22 mL/min hydrogen gas, 3.2 mL/min air and 6.4 mL/min carbon dioxide gas was supplied constantly as the main source of energy and carbon. Thus, with air, all cultivations were also supplied with nitrogen gas. Growth was monitored by taking samples automatically and analysing the cell density as optical density by measuring absorbance at 600 nm (Ultrospec 2100 pro UV/visible spectrophotometer, Biochrom Ltd., England). Growth curves of the cultivations are presented in Figure 2. Growth on ammonia and urea were comparable. Growth on nitrate or nitrogen gas was clearly slower than on ammonia or urea. Towards the end of the cultivation, the growth on nitrate was better than growth on nitrogen gas as the only source of nitrogen. There was nonetheless growth also in the cultivations in which nitrogen gas was the only source of nitrogen demonstrating that isolated bacterial strain deposited as VTT-E-193585 is capable of nitrogen fixation.

Example 4. Characterization of antibiotic susceptibility

Antibiotic susceptibility of gentamicin, kanamycin, streptomycin, tetracycline, ampicillin, ciprofloxacin, colistin and fosfomycin for the isolated bacterial strain deposited as VTT-E-193585 was analysed according to CLSI M07-A111 standard (Clinical and laboratory standards institute. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 11th ed. CLSI standard M07, 2018) with hand-made microdilution plate for ampicillin, ciprofloxacin and colistin, with VetMIC Lact-1 plate (SVA National Veterinary Institute, Uppsala, Sweden) for gentamicin, kanamycin, streptomycin and tetracycline using broth microdilution method and for fosfomycin using agar dilution method in aerobic conditions at $+35 \pm 2^\circ\text{C}$ for 48 ± 1 hours using cation-adjusted Mueller Hinton Broth -medium (LabM, LAB114, cations Mg^{2+} and Ca^{2+} added separately). *Escherichia coli* ATCC 25922 was used as quality control strain and it was incubated in aerobic conditions, at $+35 \pm 2^\circ\text{C}$ for 18 ± 2 hours. Results of antibiotic susceptibility of

strains are presented in Table 5. The isolation bacterial strain was found to be generally sensitive to antibiotics. For gentamicin, kanamycin, streptomycin and tetracycline minimum inhibitory concentration (MIC) values for VTT-E-193585 were lower or comparable to *E. coli* ATCC 25922, while for ampicillin, ciprofloxacin, colistin and fosfomycin the MIC values were higher in VTT-E-193585.

Table 5. Minimum Inhibitory Concentration (MIC, µg/ml) -values of antibiotics for VTT-E-193585 strain and *Escherichia coli* ATCC 25922

	VTT-E-193585 48 h ± 1 h	<i>E. coli</i> ATCC 25922 18 h ± 2 h
Gentamicin	0.5	0.5
Kanamycin	2	4
Streptomycin	0.5	4
Tetracycline	0.25	1
Ampicillin	16	8
Ciprofloxacin	0.06	0.008
Colistin	4	2
Fosfomycin	32	0.5

CLAIMS

1. A genetically-modified or mutated bacterial strain generated using an isolated bacterial strain deposited under number VTT-E-193585, wherein said genetically-
5 modified or mutated bacterial strain has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.
2. A culture comprising the bacterial strain according to claim 1.
- 10 3. A process for the production of biomass, said process comprising culturing the bacterial strain of claim 1, or an isolated bacterial strain deposited under number VTT-E-193585.
4. The process of claim 3, comprising culturing the bacterial strain with hydrogen
15 as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide.
5. The process according to claim 4, wherein the culture has dissolved oxygen in the culture is maintained between 5% and 10%.
- 20 6. The process according to claim 4 or 5, wherein ammonium, urea, nitrate and/or nitrogen gas is used as nitrogen source.
7. The process according to any one of claims 4 to 6, wherein the culture medium comprises minerals wherein said minerals comprises less than 1 g/L of chloride salts.
25
8. The process according to any one of claims 4 to 7, wherein the medium does not contain vitamins.
9. The processing according to any one of claims 4 to 8, wherein pH in the culture
30 is maintained between 5.5 and 8.0.
10. The process according to any one of claims 4 to 9, wherein said culture is grown at a temperature between 25°C and 40°C.
- 35 11. The process according to any one of claims 4 to 10, wherein the bacterial strain has a growth rate of 0.04-0.12 h⁻¹.

12. The process according to any one of claims 4 to 11, wherein the bacterial strain comprises the 16S ribosomal RNA set forth in SEQ ID NO:1 or a 16S ribosomal RNA having up to 20 nucleotide differences with SEQ ID NO:1, e.g. 1 to 10, such as 1 to 5, e.g. one, two, or three nucleotide differences with SEQ ID NO:1.

5

13. The process according to any one of claims 4 to 12, wherein the bacterial strain comprises:

i) a gene encoding ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) large chain having the sequence set forth in SEQ ID NO:3 or a sequence having
10 more than 93% identity to the sequence set forth in SEQ ID NO:3, and/or

ii) a gene encoding a ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) small chain having the sequence set forth in SEQ ID NO:5 or a sequence having more than 83% sequence identity to the sequence set forth in SEQ ID NO:5, and/or

15 iii) a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit alpha having the sequence set forth in SEQ ID NO:7 or a sequence having more than 70% sequence identity to the sequence set forth in SEQ ID NO:7, and/or

iv) a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit beta having the sequence set forth in SEQ ID NO:9 or a sequence having more than 77% sequence
20 identity, such as more than 80% identity, e.g. more than 90% identity to the sequence set forth in SEQ ID NO:9, and/or

v) a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit gamma having the sequence set forth in SEQ ID NO:11 or a sequence having more than 70% sequence identity to the sequence set forth in SEQ ID NO:11, and/or

25 vi) a gene encoding a NAD⁺-reducing hydrogenase HoxS subunit delta having the sequence set forth in SEQ ID NO:13 or a sequence having more than 79% sequence identity to the sequence set forth in SEQ ID NO:13, and/or

vii) a gene encoding a NiFeSe hydrogenase large subunit having the sequence set forth in SEQ ID NO:15 or a sequence having more than 84% sequence identity to
30 the sequence set forth in SEQ ID NO:15, and/or

viii) a gene encoding a NiFeSe hydrogenase small subunit having the sequence set forth in SEQ ID NO:17 or a sequence having more than 90% identity to the sequence set forth in SEQ ID NO:17, and/or

35 ix) a gene encoding an ATP synthase gamma chain atpG_1 having the sequence set forth in SEQ ID NO:19 or a sequence having more than 70% identity to the sequence set forth in SEQ ID NO:21, and/or

- xi) a gene encoding an ATP synthase subunit b atpF_1 having the sequence set forth in SEQ ID NO:23 or a sequence having more than 62% identity to the sequence set forth in SEQ ID NO:23, and/or
- xii) a gene encoding an ATP synthase subunit c, sodium ion specific atpE_1
5 having the sequence set forth in SEQ ID NO:25 or a sequence having more than 90% identity to the sequence set forth in SEQ ID NO:25, and/or
- xiii) a gene encoding an ATP synthase subunit a atpB_1 having the sequence set forth in SEQ ID NO:27 or a sequence having more than 80% identity to the sequence set forth in SEQ ID NO:27, and/or
- xiv) a gene encoding an ATP synthase epsilon chain atpC_1 having the
10 sequence set forth in SEQ ID NO:29 or a sequence having more than 71% identity to the sequence set forth in SEQ ID NO:31, and/or
- xvi) a gene encoding an ATP synthase subunit beta atpD_2 having the sequence set forth in SEQ ID NO:33 or a sequence having more than 97% identity to
15 the sequence set forth in SEQ ID NO:33, and/or
- xvii) a gene encoding an ATP synthase gamma chain atpG_2 having the sequence set forth in SEQ ID NO:35 or a sequence having more than 86% identity to the sequence set forth in SEQ ID NO:35, and/or
- xviii) a gene encoding an ATP synthase subunit alpha atpA_2 having the
20 sequence set forth in SEQ ID NO:37 or a sequence having more than 98%, to the sequence set forth in SEQ ID NO:37, and/or
- xix) a gene encoding an ATP synthase subunit delta atpH having the sequence set forth in SEQ ID NO:39 or a sequence having more than 85% identity to the sequence set forth in SEQ ID NO:39, and/or
- xx) a gene encoding an ATP synthase subunit b atpF_2 having the sequence
25 set forth in SEQ ID NO:41 or a sequence having more than 87% identity to the sequence set forth in SEQ ID NO:41, and/or
- xxi) a gene encoding an ATP synthase subunit b' atpG_3 having the sequence set forth in SEQ ID NO:43 or a sequence having more than 81% identity to the sequence
30 set forth in SEQ ID NO:43, and/or
- xxii) a gene encoding ATP synthase subunit c atpE_2 having the sequence set forth in SEQ ID NO:45 or a sequence having more than 98% to the sequence set forth in SEQ ID NO:45, and/or
- xxiii) a gene encoding an ATP synthase subunit a atpB_2 having the sequence
35 set forth in SEQ ID NO:47 or a sequence having more than 92% identity to the sequence set forth in SEQ ID NO:47, and/or

xiv) a gene encoding an ATP synthase protein I atpI having the sequence set forth in SEQ ID NO:49 or a sequence having more than 60% identity to the sequence set forth in SEQ ID NO:49, and/or

5 xv) a gene encoding a nitrogenase molybdenum-iron protein alpha chain nifD_1 having the sequence set forth in SEQ ID NO:51 or a sequence having more than 60% identity to the sequence set forth in SEQ ID NO:51, and/or

xvi) a gene encoding nitrogenase molybdenum-iron protein alpha chain nifD_2 having the sequence set forth in SEQ ID NO:53 or a sequence having more than 60% identity to the sequence set forth in SEQ ID NO:53, and/or

10 xvii) a gene encoding a nitrogenase molybdenum-iron protein beta chain nifK_1 having the sequence set forth in SEQ ID NO:55 or a sequence having more than 87% identity to the sequence set forth in SEQ ID NO:55, and/or

xviii) a gene encoding a nitrogenase molybdenum-iron protein beta chain nifK_2 having the sequence set forth in SEQ ID NO:57 or a sequence having more than 15 95% identity to the sequence set forth in SEQ ID NO:57, and/or

xxix) a gene encoding a nitrogenase iron protein nifH having the sequence set forth in SEQ ID NO:59 or a sequence having more than 98.5% sequence identity to the sequence set forth in SEQ ID NO:59, wherein the strain preferably comprises:

- the genes described in iii), iv), v) and vi),
- 20 • the genes described in vii) and viii), or
- the genes described in iii), iv), v), vi), vii) and viii).

14. The process according to any one of claims 4 to 13, comprising the further step of harvesting biomass produced during the culture, optionally comprising a further step 25 of drying the biomass.

15. A process for the production of protein, comprising performing the process according to claim 14 and a further step of isolating protein from said biomass, wherein the process results in a protein fraction and a fraction comprising non-protein 30 components.

16. The process according to claim 14 or 15, comprising the further step of producing a food or feed product from said biomass, from said protein fraction or from said fraction comprising non-protein components.