

STRAINS 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
15 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
20 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. Andersen et al. (1979) Biochim Biophys Acta 585:1-11 describes mutant strains of *Alcaligenes eutrophus*, a hydrogen bacterium that grows readily under
25 heterotrophic and autotrophic conditions. Mutants having altered ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) activity were characterised. Ohmiya et al. (2003) J. Biosci. Bioeng. 95:549-561 reviews the application of microbial genes to recalcitrant biomass utilization. Yu Jian et al. (2013) Int J Hydrogen Ener 38:8683-8690 describes carbon dioxide fixation by a hydrogen-oxidizing bacterial isolate. A high
30 energy efficiency of 50% was measured under a moderate oxygen concentration (10 mol%).

However, various chemoautotrophic microorganisms have different properties in terms of growth rate, yield, biomass composition as well as properties related to being used
35 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.

5

SUMMARY OF THE INVENTION

In a first main aspect, the invention relates to an isolated bacterial strain VTT-E-193585 or a derivative thereof.

- 10 In further aspects, the invention relates to a culture comprising the bacterial strain of the invention or derivative thereof. 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 a derivative thereof.
- 15 In a further aspect, the invention relates to a process for the production of biomass and/or protein, said process comprising culturing a bacterial strain of the genus *Xanthobacter* in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide.
- 20 In further main aspects, the invention relates to bulk protein, biomass or non-protein cellular or chemical components obtained or obtainable by the process of the invention, and to a food or feed product obtained or obtainable by a process of the invention.

BRIEF DESCRIPTION OF THE FIGURES

- 25 **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
- 30 nitrogen sources.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

- When used herein, the term "isolated", e.g. in the context of a strain, means isolated
- 35 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 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 to the reference strain.

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

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.

Aspects and embodiments of the invention

In a first main aspect, the invention relates to an isolated bacterial strain VTT-E-193585 or a derivative thereof.

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.

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.

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.

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.

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

```
CTTGAGAGTTTGATCCTGGCTCAGAGCGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGA
GCGCCCAGCAATGGGAGCGGCAGACGGGTGAGTAACGCGTGGGGATGTGCCCAATGGTACG
20 GAATAACCCAGGGAACTTGGACTAATACCGTATGAGCCCTTCGGGGGAAAGATTTATCGCCA
TTGGATCAACCCGCGTCTGATTAGCTAGTTGGTGGGGTAACGGCCCCACCAAGGCGACGATCA
GTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGTG
TGATGAAGGCCTTAGGGTTGTAAAGCACTTCGCCGGTGAAGATAATGACGGTAACCGGAGA
25 AGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCG
GAATCACTGGGCGTAAAGCGCACGTAGGCGGATCGTTAAGTCAGGGGTGAAATCCTGGAGCT
CAACTCCAGAACTGCCCTTGATACTGGCGACCTTGAGTTCGAGAGAGGTTGGTGAACTGCG
AGTGTAGAGGTGAAATTCGTAGATATTCGCAAGAACACCAAGTGGCGAAGGCGGCCAACTGGC
TCGATACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTC
30 CACGCCGTAAACGATGGATGCTAGCCGTTGGGCAGCTTGCTGTTCAAGTGGCGCAGCTAACGC
ATTAAGCATCCCGCCTGGGGAGTACGGTCGCAAGATTAAACTCAAAGGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTGAAGCAACGCGCAGAACCTTACCAGCCTTT
GACATGGCAGGACGATTTCCAGAGATGGATCTCTCCAGCAATGGACCTGCACACAGGTGCT
GCATGGCTGTCGTGAGCTCGTGTGCTGAGATGTTGGGTTAAGTCCCACGAGCGCAACCC
35 TCGCCTCTAGTTGCCAGCATTAGTTGGGCACTCTAGAGGGACTGCCGGTGATAAGCCGAGA
GGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACA
ATGGTGGTGACAGTGGGATGCGAAAGGGCGACCTCTAGCAAATCTCCAAAAGCCATCTCAGT
```

TCGGATTGTACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGTGGATCAGCAT
GCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCTT
TACCCGAAGGCGCTGCGCTAACCCGCAAGGGAGGCAGGCGACCACGGTAGGGTCAGCGACT
GGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCTGCGGCTGGATCACCTCCTTT

5

In a further aspect, the invention relates to a culture comprising the bacterial strain of the invention or derivative thereof. 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.
10 200,000 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 derivative thereof. In one embodiment, the process is for the production of biomass.
15 In another embodiment, the process is for the production of protein. 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. In a further embodiment, the process is for the production of biomass and comprises culturing the strain in continuous culture with
20 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.

In a further main aspect, the invention relates to a process for the production of biomass and/or protein, said process comprising culturing a bacterial strain of the genus *Xanthobacter* in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide. In one embodiment, the process is for the production of biomass. In another embodiment, the process is for the production of protein. Various further embodiments of the process are
25 described herein below.
30

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
35 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* 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%.

5

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%.

10

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

15

between 5% and 10%.

In a preferred embodiment, only gases and minerals are provided to the culture and the gas provided comprising H_2 , CO_2 and O_2 , wherein the percentage of H_2 is between 40% and 70%, the percentage of CO_2 is between 18% and 28% and the percentage of O_2 is

20

between 12% and 22%.

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

25

nitrate, e.g. potassium nitrate. In other embodiments, nitrogen gas (N_2) is provided as nitrogen source. In a preferred embodiment, the nitrogen source is ammonium hydroxide or an ammonium salt, such as ammonium sulphate.

In one embodiment, the nitrogen source provided is ammonium hydroxide at a

30

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.

Typically, the process of the invention includes the addition of minerals, such as minerals containing ammonium, phosphate, potassium, sodium, vanadium, iron, sulphate,

35

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 (NH_4Cl), ammonium sulphate ($(\text{NH}_4)_2\text{SO}_4$)), nitrate (e.g., potassium nitrate (KNO_3)), urea or an organic nitrogen source; phosphate (e.g., disodium phosphate (Na_2HPO_4), potassium phosphate (KH_2PO_4), phosphoric acid (H_3PO_4), potassium dithiophosphate ($\text{K}_3\text{PS}_2\text{O}_2$), potassium orthophosphate (K_3PO_4), disodium phosphate ($\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$), dipotassium phosphate (K_2HPO_4) or monopotassium phosphate (KH_2PO_4); sulphate; yeast extract; chelated iron (chelated e.g. with EDTA or citric acid); potassium (e.g., potassium phosphate (KH_2PO_4), potassium nitrate (KNO_3), potassium iodide (KI), potassium bromide (KBr)); and other inorganic salts, minerals, and trace nutrients (e.g., sodium chloride (NaCl), magnesium sulphate ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$) or magnesium chloride (MgCl_2), calcium chloride (CaCl_2), 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 (CuSO_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.

- 5 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
10 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
15 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.

20 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
25 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
30 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
35 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.

5

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.

10

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.

15

In one embodiment of the process of the invention, the bacterial strain has a growth 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.

20

Xanthobacter is a genus of Gram-negative bacteria from the *Xanthobacteraceae* family.

In one embodiment, the *Xanthobacter* strain used in the process of the invention is a strain which uses the Calvin Benson Bassham pathway to convert carbon dioxide into organic compounds, e.g. glucose, essential for living organisms.

25

In one embodiment, the *Xanthobacter* strain used in the process of the invention is a strain which uses NiFeSe-hydrogenases for converting hydrogen (H₂) into cellular energy equivalents.

30

In one embodiment, the *Xanthobacter* strain used in the process of the invention is a strain which uses NAD⁺-reducing hydrogenases for converting hydrogen (H₂) into cellular energy equivalents.

35

In one embodiment, the *Xanthobacter* strain used in the process of the invention capable of nitrogen fixation.

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.

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 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 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 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.

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 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 bisphosphate carboxylase large chain:

ATGGGTGCCGAAGCAACCGTCGGGCAGATCACGGACGCCAAGAAGAGATACGCCGCCGGCG
TGCTGAAGTACGCCAGATGGGCTACTGGAACGGCGACTACGTTCCCAAGGACACCGACCTC
CTGGCGGTGTTCCGCATCACCCCCAGGCGGGCGTGGACCCGGTGGAAGCCGCCGCGGCGG
TCGCCGGCGAAAGCTCCACCGCTACCTGGACCGTGGTGTGGACCGACCGGCTACCGCCGC
CGACGTCTACCGCGCCAAGGCCTACAAGGTGGAGCCGGTGCCGGGCCAGGAAGGCCAGTAT

TTCTGCTACATCGCCTATGATCTCGATTGTTCGAGGAAGGCTCCATCGCCAACTCACGGCG
TCGATCATCGGCAACGTCTTCTCCTTCAAGCCGCTGAAGGCGGCGCGGCTGGAGGACATGCG
GCTTCCCGTCGCCTATGTGAAGACCTTCCGCGGCCCCGCCACCGGCATCGTGGTCGAGCGCG
AGCGCCTGGACAAGTTCGGCCGCCCCCTTCTGGGCGCCACCACCAAGCCGAAGCTTGGCCTC
5 TCGGGCAAGAATTACGGCCGCGTGGTCTATGAGGCCCTCAAGGGCGGCCTCGACTTCGTGAA
GGACGACGAGAACATCAACTCGCAGCCCTTCATGCACTGGCGCGATCGCTTCTCTATTGCAT
GGAGGCCGTCAACAAGGCCAGGCCGAGACCGGCGAGGTGAAGGGGCACTATCTCAACATC
ACCGCCGGGACCATGGAGGAGATGTACCGCCGCGCCGAGTTCGCCAAGGAACTGGGCTCCG
TGGTGGTGTATGGTGGATCTCATCATCGGCTGGACCGCCATCCAGTCCATGTCCAACCTGGTGC
10 CGCGAGAACGACATGATCCTGCACATGCACCGTGCGGGCCATGGCACCTACACGCGCCAGAA
GAGCCACGGCGTCTCCTTCCGCGTCATCGCCAAGTGGCTGCGGCTCGCCGGCGTCGACCACC
TGCACACCGGCACCGCCGTGGGCAAGCTGGAAGGCGACCCCATGACCGTGCAGGGCTTCTA
CAATGTCTGCCGCGAGACGACGACGACGACGAGGACCTCACCCGCGGCCTGTTCTTCGAGCAGG
ACTGGGGCGGCATCCGCAAGGTGATGCCGGTGGCCTCCGGCGGCATCCATGCGGGCCAGAT
15 GCACCAGCTCATCGACCTGTTCCGGCGAGGACGTGGTGTCCAGTTCGGCGGGCGGCACCATCG
GCCACCCGGACGGCATCCAGGCCGGCGCCACCGCCAACCGCGTGGCGCTGGAAACCATGAT
CCTCGCCCGCAACGAGGGCCGCGACATCAGGAACGAGGGCCCGGAAATCCTGGTGGAAAGCC
GCCAATGGTGCCGTCCGCTGCGCGCGGCGCTCGATACCTGGGGCGAGGTGACCTTCAACTA
CGCCTCCACCGACACGTCCGATTACGTGCCACCGCGTCCGTGCGCTGA

20

SEQ ID NO: 3:

Amino acid sequence of Ribulose biphosphate carboxylase large chain

MGAEATVGQITDAKKRYAAGVLKYAQMGYWNGDYVPKDTDLLAVFRITPQAGVDPVEAAAAVA
GESSTATWTVVWTDRLTAADVYRAKAYKVEPVPGQEGQYFCYIAYDLDFEEGSIANLTASIIGN
25 VFSFKPLKAARLEDMRLPVAYVKTRGPPTGIVVERERLDKFRPLLGATTKPKLGLSGKNYGRVV
YEALKGGDLDFVKDDENINSQPFMHWDRFLYCM EAVNKAQAETGEVKGHYLNITAGTMEEMYRR
AEFAKELGSVVVMVDLIIGWTAIQSMSNWCRENDMILMHMRAGHGTYTRQKSHGVSFRVIAKW
LRLAGVDHLHTGTAVGKLEGPMTVQGIFYNCRETTTQQDLTRGLFFEQDWGGIRKVMVASG
GIHAGQMHQLIDLFGEDVVLQFGGGTIGHPDGIQAGATANRVALETMILARNEGRDIRNEGPEIL
30 VEAAKWCRPLRAALDTWGEVTFNYASTDTSYVPTASVA

SEQ ID NO: 4:

Nucleotide sequence of Ribulose biphosphate carboxylase small chain:

ATGCGCATACCCAAGGCTCCTTCTCCTTCTGCCGGACCTACCGACACGACATCAAGGCC
35 CAGGTGCAATATTGCCTGGACCAGGGCTGGGCGGTCTCGGTGGAGCACACCGACGATCCCCA
CCCGCGCAACACCTATTGGGAGATGTGGGGCCCGCCCATGTTTCGATCTGCGCGACGCGGCC
GGCGTCTTCGGCGAGATCGAAGCCTGCCGGGCCGCCAATCCCGAGCATTATGTGCGGGTGAA

CGCCTTCGATTCCAGCCGCGGATGGGAGACGATCCGCCTGTCCCTTCATCGTTCAGCGGCCCA
CCGTGGAAGAGGGCTTCCGCCTCGACCGCACCGAAGGCAAGGGCCGCAACCAGAGCTACGC
CATGCGCTACCGGGCGCAGTTCGCGCCGCGCTGA

5 SEQ ID NO: 5:

Amino acid sequence of Ribulose biphosphate carboxylase small chain:

MRITQGSFSFLPDLTDQIKAQVQYCLDQGWAVSVEHTDDPHPRNTYWEMWGPPMFDLRDAAG
VFGEIEACRAANPEHYVRVNAFDSSRGWETIRLSFIVQRPTVEEGFRLDRTEGKGRNQSYAMRYR
AQFAPR

10

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. 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 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.

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 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

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.

5

SEQ ID NO:6:

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

ATGATGCCATCTGAGCCGCACGGCGCGGGCATGCCGCCCCACGGGAAGCGGCCGCGGTTC
CCACCCCCCAGGAGGTGAGCGCGGTGGTGGCCGAGGTGGTCGCGGATGCCGTGGCATCGGT
10 GGGCGGCGCACGCACCCGGCTCATGGACATCGTCCAGCTGGCCCAGCAGCGTCTCGGCCAT
CTCTCCGAAGAGACCATGGCGGCCATTGCCGCGCGGCTCGCCATTCCGCCGGTGGAAGTGG
CGGACATGGTGTCTTCTACGCCTTCTGAACCGCGCGCCCAAGGGCCGCTACCACATCCGC
CTGTGCGCGAGCCCCATCTCGCTGATGAAGGGCGCCGAGGCGGTGGTGCCGCCTTCTGCCA
GATCCTCGGCATCGCCATGGGCGAGACCTCGCAGGATGGCGACTTCACCCTGGAATGGACCA
15 ACGACATCGGCATGGCCGACCAGGAGCCGCGCCCTCGTCAACGGCACGGTGATGACGCA
GCTCGCGCCCGGCGATGCGGCCATCATCGTGGCCGGCTGCGGGCCCCATCACGCGCCCAAT
GCCCTGCCGCTGTTCCCTGGAGCCGGCGTGGCCGGCTCCGGCCTGCCCCATGCCCGGATCC
GCCCCAGCCTGGTGATGCCGGGACAGCTTCTGTTCCGCGAGGACCACACGACGCCGGGCGC
CGGCATCAAGGCGGCACTCGCCCTACCCCGGACGAAGTGGTGCAGAAGGTCTCCGCCGCG
20 CGCCTGCGCGGGCGGGGTGGCGCCGGCTTTCCACCGGTCTCAAATGGAAGCTCTGCCGCC
AGTCGCCCCGCCACCAACCGCCATGTGATCTGCAATGCGGACGAGGGCGAGCCCGGCACCTTC
AAGGATCGCGTGCTGCTCACGCAGGCGCCGCACCTCATGTTGACGGCATGACCATCGCCGG
CTACGCCTTGGGGGCGCGGGAGGGCGTGGTCTATCTGCGCGGCGAGTACGCCTATCTGTGG
GAGCCTCTGCATGCGGTCTGCGCGAGCGCTATGGGCTCGGGCTCGCCGGCGCGAACATCC
25 TGGGACACGCGGGCTTCGACTTCGACATCCGCATCCAGCTGGGCGCCGGCGCCTATATCTGC
GGCGAGGAATCCGCGCTGGTGAATCGCTGGAAGGCAAGCGCGGCTCGCCCCGCGACCGCC
CCCCCTTCCCCACCGTGCGCGGCCATCTCCAGCAGCCCACCGCCGTGGACAATGTGGAGACC
TTCGCCTGCGCCGCCCGCATCCTGGAGGATGGCGTGGAGGCGTTGCGGGGCATCGGCACGC
CCGAATCCGCCGGCACGAAGCTCCTCTCGGTGTCGGGCGATTGCCCCGCGCCCCGGCGTGTAT
30 GAGGTGCCCTTCGGCCTCACGGTGAACGCGCTGCTCGACCTTGTGCGCGCGCCGGACGCCG
CCTTCGTGCAGATGGGTGGGCCGTCCGGCCAATGCGTGGCGCCGAAGGATTACGGCCGCCG
CATCGCCTTCGAGGACCTGCCCACCGGCGGCTCGGTGATGGTGTTCGGCCCCGGGGCGCGAC
GTGCTCGCCATGGTGCGCGAGTTCGCGGATTTCTTCGCCGGCGAATCCTGCGGCTGGTGAC
GCCCTGCCGGGTGGGCACCACTTGCTCAAGGAAGAGCTGGACAAGCTCCTCGCCAACCGCG
35 CCACCCTCGCCGACATCCGCGCGCTGGAGACCCTGGCCACGACCGTCTCCCGCACCAGCCGC
TGCGGCCTCGGCCAGACGGCGCCCAACCCCATCTTTCCACCATGCGCAACCTGCCGGAAGC

CTATGAGGCGAGGCTGAGGCCCCGAAGACTTCCTGCCCTGGGCCTCGCTCGACGAGGCGCTG
AAGCCCGCCATCGTCATCCAGGGCCGCGCGCCCGTGCCGGAGGAAGAGGCATGA

SEQ ID NO: 7:

- 5 Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit alpha:
MMPSEPHGAGMPPPREAAVPTPQEVSAVVAEVDADAVASVGGARTRLMDIVQLAQQLGHLSE
ETMAAIAARLAIPPEVADMVSFYAFLNRAPKGRYHIRLSRSPISLMKGAEAVAAAFQCILGIAMG
ETSQDGDFTLEWTNDIGMADQEPAALVNGTVMTQLAPGDAAIIVGRLRAHHAPNALPLFPGAGV
AGSGLPHARIRPSLVMPGQLLFREDHTTPGAGIKAALALTPDEVVQKVSAARLRGRGGAGFPTGL
10 KWKLCRQSPATTRHVICNADEGEPTFKDRVLLTQAPHLMFDGMTIAGYALGAREGVVYLRGEY
AYLWEPLHAVLRERYGLGLAGANILGHAGFDFFDIRIQLGAGAYICGEESALVESLEGKRGSPRDRP
PFPTVRGHLQQPTAVDNVETFACAARILEDGVEAFAGIGTPESAGTKLLSVSGDCPRPGVYEVFPFG
LTVNALLDLVGAPDAAVQMGGPSGQCVPKDYGRRIAFEDLPTGGSVMVFGPRDVLAMVREF
ADFFAGESCGWCTPCRVGTTLLKEELDKLLANRATLADIRALETATTVSRTSRCGLGQTAPNPIL
15 STMARNLPEAYEARLRPEDFLPWASLDEALKPAIVIQGRAPVPEEEA

SEQ ID NO: 8:

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

- ATGAGCCGGGGATCCCCGATGCCGGAAAGACCGCACCATGAGCGCCACCGACGGCACCA
20 CCGCCCCCGCAAGATCGTCATCGATCCGGTGACCCGCGTGGAGGGCCACGGCAAGGTCAC
CATCCGCCTGGATGAAGCCGGCGCGGTGGAGGATGCGCGTTTCCACATCGTGAGTTCGCGC
GGCTTCGAGCGGTTTCATCCAGGGCCGGATGTAAGGAGTGCCCTTATCATCCAGCGGCT
GTGCGGCATCTGCCCCGTGAGCCACCATCTGGCGGCGGCGAAAGCCATGGACCAGGTGGCG
GGCGTGAGCCGCTACCGCCACCGCCGAGAACTGCGCCGGCTGATGCATTATGGGCAGG
25 TGCTGCAATCCAACGCTTTCACATCTTCCACCTCGCCTCGCCGACCTCCTGTTGCGGCTTCG
ACGCGCCGGCCGAGCAGCGCAACATCATCGCCGTGCTCCAGCGTTATCCGGAGATCGGCAAA
TGGGCGATCTTCATCAGGAAGTTCGGCCAGGAGGTCATCAAGGCCACCGGCGGGCGCAAGA
TCCATCCACAGCGCCATTCCCGGCGGGGTCAACCAGAACCTCGCCGTGGAGGACCGCGAC
GCCCTGCGCGCCAAGGTGGGCGAGATCATCAGCTGGTGATGGCGGCGCTGGACCATCACA
30 AGGCCTATGTGGCGGAAACCGGGCGCTGCATGACAGCTTCGCCGCCTTCCCCTCCGCCTTC
ATGAGCCTCGTGGGGCCGGATGGCGGCATGGACCTTTATGACGGCACCTGCGGGTGATCG
ATGCCGAGGGCGCCCCCTCATCGAAGGCGCGCCGCCCGCCTCCTACCGCGACCACCTCATC
GAGGAGGTGCGGCCCTGGAGCTATCTGAAATTCCTCCATCTGCGCGCCTTCGGCCGCGACGA
TGGCTGGTATCGGGTCGGCCCCCTCGCCAGGTCAATTGCGCCGCGTCCATCGACACGCCCC
35 GCGCCGAGGCGGCGGGGCGGACTTCATGGCCGAGGGCGGCGGCAAGCCGGTGATGCCA
CCCTCGCTTATCACTGGGCGCGGCTCATCGTGCTGGTCCATTGCGCGGAGAAGATCGAACAG
CTGCTGTTGACGACGACCTGCAAGGCTGCGATCTGCGTGCGGAGGGCACCCGGCGCGGGG

AAGGCGTCGCCTGGATCGAGGCGCCGCGCGGCACCCTCATCCACCATTACGAGGTGGACGA
GAACGACCAGGTGCGCCGCGCCAACCTCATCGTCTCCACCACCCACAATAACGAGGCCATGA
ACCGCGCCGTGCGGCAGGTGGCGAAGACGGACCTTTCCGGTCGCGAGATCACCGAAGGGCT
GCTGAACCATATCGAGGTGGCCATCCGCGCCTTCGACCCCTGCCTGTCCTGCGCCACCCATG
5 CGCTGGGCCAGATGCCGCTGATCGTGACGCTTGAAGATGCCTCCGGCGCAGAGATCGCCCCG
CGGAGTGAAGGAATGA

SEQ ID NO: 9:

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

10 MSRGSPDAGKDRTMSATDGTTPRKIVDPVTRVEGHGKVTIRLDEAGAVEDARFHIVEFRGFER
FIQGRMYWEVPLIIQRLCGICPVSHHLAAKAMDQVAGVDRVPPTAEKLRRMLMHYGQVLQSNAL
HIFHLASPDLLFGFDAPAEQRNIIAVLQRYPEIGKWAIFIRKFGQEVIKATGGRKIHPTSAIPGGVN
QNLAVEDRDALRAKVGEIISWCMAALDHHKAYVAENRALHDSFAAFPSAFMSLVGPDGGMDLY
DGTLRVIDAEGAPLIEGAPPASYRDHLIEEVRPWSYLFPHLRAFRDDGWYRVGPLAQVNCAAS
15 IDTPRAEAARRDFMAEGGGKPVHATLAYHWARLIVLVHCAEKIEQLLFDDDLQGCDLRAEGTRRG
EGVAWIEAPRGTLIHYYEVDENDQVRRANLIVSTTHNNEAMNRAVRQVAKTDLSGREITEGLLN
HIEVAIRAFDPCLSCATHALQMPLIVTLEDASGAIEIARGVKE

SEQ ID NO: 10:

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

ATGAGCGAGACCCCTTCACCTTTACCGTGGACGGCATCGCGGTCCCGGCCACCCCGGCCA
GAGCGTCATCGAGGCGTGCATGCGGCGGGCATCTATATCCCGCGCCTGTGCCACCACCCG
GACCTGCCGCCGCGGGGCCATTGCCGGGTGTGCACCTGCATCATCGACGGGCGGCCGCCA
GCGCCTGCACCATGCCCCGCCAGGGGCATGGTGGTGGAGAACGAGACGCCCGCTTTGCT
25 GGCGGAGCGGCGCACGCTGATCGAGATGCTGTTGCGGAAGGCAACCATTTCTGCCAGTTCT
GCGAGGCGAGCGGCGATTGCGAATTGCAGGCGCTGGGCTACCTGTTGGCATGGTGGCCCC
GCCCTTCCCCCATCTGTGGCCGAAGCGGCCGGTGGATGCCAGCCATCCGGATATCTATATCG
ACCACAATCGCTGCATCCTGTGCTCGCGCTGCGTGCGCGCCTCGCGCACCTGGACGGCAAG
TCCGTGTTGGCTTCGAGGGGCGCGGCATCGAGATGCATCTGGCGGTGACCGGCGGGCACC
30 TGGACGACAGCGCCATCGCCGCCGCCGACAGGGCGGTTGAGATGTGCCCGGTGGGCTGCAT
CGTCCTCAAGCGCACCGGCTACCGCACGCCCTATGGCCGGCGGCGCTACGACGCCGCGCCC
ATCGGCTCCGACATACCGCCCCGCGCGGCGGCGCGGAAGGACTGA

SEQ ID NO: 11:

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

MSETPFTFTVDGIAPATPGQSVIEACDAAGIYIPRLCHHPDLPPAGHCRVCTCIIDGRPASACTM
PAARGMVVENETPALLAERRTLIEMLFAGNHFCQFCEASGDCELQALGYLFGMVAPPFPHLWPK

RPVDASHPDIIYIDHNRCLCSRCVRSRTLDGKSVFGFEGRGIEHMLAVTGGHLDSDAIAAADRA
VEMCPVGCIVLKRTGYRTPYGRRRYDAAPIGSDITARRGGAKD

SEQ ID NO: 12:

5 Nucleotide sequence of NAD⁺-reducing hydrogenase HoxS subunit delta:
ATGGCCAAGCCCAAACTCGCCACCTGCGCGCTGGCCGGCTGCTTCGGCTGCCACATGTCCTT
CCTGGACATGGACGAGCGCATCGTCGAGCTCATCGACCTGGTGGACCTCGACGTCTCGCCCC
TCGACGACAAGAAAACTTCACCGGCATGGTGGAAATCGGCCTGGTGGGAAGGCGGCTGCGC
CGACGAGCGCCATGTGAAGGTGCTGCGCGAGTTCGCGAGAAATCCCGCATCCTGGTGGCG
10 GTGGGCGCCTGCGCCATCACCGGCGGCATCCCGGCATTGCGCAACCTCGCCGGCCTCGACG
AATGCCTGAGGGAAGCCTACCTCACCGGCCCCACGGTGGGAAGGCGGCGGGCTCATTCCCAAC
GACCCGGAGCTGCCGCTGCTGCTGGACAAGGTCTATCCGGTGCAGGACTTCGTGAAGATCGA
CCATTTCTGCGCCGGCTGCGCCCTCGGCCGACCCATCTGGGCGGCTCTGAAGGCGCTGC
TGACCGGCACCGAGCCGCATCTGCCCTACCCGCTTTTCAAGTACGAATGA

15

SEQ ID NO: 13:

Amino acid sequence of NAD⁺-reducing hydrogenase HoxS subunit delta:
MAKPKLATCALAGCFGCHMSFLDMDERIVELIDLVDLVSPLDDKKNFTGMVEIGLVEGGCADER
HVKVLREFREKSRILVAVGACAITGGIPALRNLAGLDECLREAYLTGPTVEGGGLIPNDPELPLLLD
20 KVYPVQDFVKIDHFLPGCPPSADAIWAALKALLTGTEPHLPYPLFKYE

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.
25 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: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
30 SEQ ID NO:17.

35

SEQ ID NO: 14:

Nucleotide sequence of Periplasmic [NiFeSe] hydrogenase large subunit:

TCCAGACCCGGGCAACATTGCTCCATGTGCTGGGCACCCTGGCCGGCCGCTGGCCCCATACC
CTCGCGCTCCAGCCCGGCGGGGTGACCCGAAGCGCCGACCAGCACGACCGCATGCGCCTGC
TCGCGACGCTGAAGGCGGTGCGGGCGGCGCTGGAAGAGACCTTGTTTCGGCGCGCCTTTGGA
AGAGGTGGCGGCCCTGGACGGCGCCGCCGCGCTGGAGGCCTGGCGCGCCAACGGCCCCGGA
5 AGGGGATTTCCGCCTGTTCTCTGGAGATCGCCGCCGACCTGGAGCTGGACCGGCTCGGCCGC
GCGCACGACCGCTTTCTCTCCTTCGGCGCCTACGCCCAGGACGAGGGGCGCCTTTATGGCGC
CGGCACCTTCGAGGCCGGGACGGCGGGAGGGCTCGATCCCAACGCCATCACCGAGGACCAC
GCCTTCGCCCCGCATGGAGGACCGCGCGGCGCCCCATGCGCCCTTTGACGGCTCCACCTTCCC
CGATGCCGACGACACCGAGGGCTACACCTGGTGCAAGGCGCCGCGCCTTGCCGGCCTGCCC
10 TTCGAGACCGGCGCCTTCGCCCGGCAGGTGGTGGCGGGCCATCCGCTCGCCCGGGACCTCG
TGACGCGGGAAGGCGGCACTGTGCGCAGCCGCGTGGTGGCCGGCTGCTGGAAACCGCGC
GCACCCTGATCGCCATGGAGGGCTGGGTGAAGGAACTGCGGCCCCGAAGGGCCCTGGTGC
CCAGGGCCACCTGCCCCAGGAAGGCCGCGCCTTCGGCCTCACCGAGGCGGCGCGGGGCG
GCTCGGCCACTGGATGGTGGTGGAGAAGGGCCGCATTGCCCCTACCAGATCATCGCCCCCA
15 CCACCTGGAATTCTCCCCCGCGACGGCGCGGGCCTGCCCGGCCGCTGGAGACGGCCCT
GGTGGGCGCGCCGTGCGGCAGGGAGAGACGACGCCCGTGAGCGTGACGACATCGTGCG
CTCCTTCGACCCGTGCATGGTCTGCACTGTGCATTGA

SEQ ID NO: 15:

20 Amino acid sequence of Periplasmic [NiFeSe] hydrogenase large subunit:
MSAETRRLLVGPFNVEGDLEVRDLVDGRVQQAFVSSPLFRGFERILEGRDPRDALVIAPRICGI
CSVSQSHAAALALAGLQGIAPTHDGRIATNLIVAENVADHLTHFHVFFMPDFARAVYEDRPWFA
QAARRFKANQGVSVRRALQTRATLLHVLGTLAGRWPHTLALQPGGVTRSADQHDRMRLLATLKA
VRAALEETLFGAPLEEVAALDGAHAVEAWRANGPEGDFRLFLEIAADLELDRLGRAHDRFLSFGAY
25 AQDEGRLYGAGTFEAGTAGGLDPNAITEDHAFARMEDRAAPHAPFDGSTFPDADDTEGYTWCK
APRLAGLPFETGAFARQVVAGHPLARDLVTREGGTVRSRVVGRLLETARTLIAMEGWVKELRPEG
PWCAQGHLPQEGRAFLTEAARGALGHWMVVEKGRIARYQIIAPTTWNFSPRDGAGLPGPLETA
LVGAPVRQGETTPVSVQHIVRSFDPVMVCTVH

30 SEQ ID NO: 16:

Nucleotide sequence of Periplasmic [NiFeSe] hydrogenase small subunit
ACGGGGGAGGAAGCCCGCGCCATCTTCGACGCCATCCTTGCCGGCGTTATCGTCCTCGACGC
CCTGTGCGTGGAAGGCGCGCTGCTGCGCGGGCCGAACGGCACCGGGCGCTTCATGTGCTG
GCGGGCACGGACACCCCCACCATCGACTGGGCGCGGCAGCTCGCCGGCATGGCGCGCCACG
35 TGGTGGCGGTGGGCACCTGCGCCGCTATGGGGGCGTGACGGCGGCGGGCATCAACCCAC
CGATGCCTGCGGCCTCCAGTTCGACGGACGCCGAAGGGTGGGGCGCTGGGGGCGGACTTC
CGCTCCCGCTCGGGGCTTCGGTCATCAATGTGGCCGGCTGCCCCACCCATCCCAACTGGGT

GACGGAAACCTGATGCTGCTCGCCTGCGGCTGCTGGGCGAGGCCGACCTCGACGTCTATG
 GCCGCCCCGCGCTTCTATGCGGACCTGCTGGTGCATCACGGCTGCCGCGCAACGAATACTAT
 GAATACAAGGCGAGCGCCGAGAAGATGAGCGACCTCGGCTGCATGATGGAGCATCTGGGCT
 GCCTCGGCACCCAGGCCACGCCGACTGCAACACGCGCCTTTGGAATGGCGAGGGCTCGTG
 5 CACCCGCGGCGGCTATGCCTGCATCAACTGCACGGCGCCGGAATTCGAGGAGCCGGGCCAC
 GCCTTCCTGGAGACGCCAAGATCGGCGGCATCCCCATCGGCCTGCCACCGACATGCCCAA
 GGCCTGGTTCATCGCCTTGTCTCCCTCGCCAAGGCGGCGACGCCGGAGCGGCTGCGCAAG
 AACGCGGTGTCCGACCATGTGGTCACGCCGCCCGCCGTCAAGGACATCAAGCGGCGATGA

10 SEQ ID NO: 17:

Amino acid sequence of Periplasmic [NiFeSe] hydrogenase small subunit

MSTPFSVLWLQSGGCGGCTMSLLCAEPDLATTLDAAGIGFLWHPALSEETGEEARAIFDAILAG
 VIVLDALCVEGALLRGPNGTGRFHVLAGTDTPTIDWARQLAGMARHVAVGTCAAYGGVTAAGI
 NPTDACGLQFDGRRKGGALGADFRSRSGLPVINVAGCPHPNWWVTETMLLACGLLGEADLDVY
 15 GRPRFYADLLVHHGCPRNEYEYKASAEKMSDLGCMMEHLGCLGTQAHADCNTRLWNNEGSC
 RGGYACINCTAPEFEPEGHAFLETPKIGGIPIGLPTDMPKAWFIALSSLAKAATPERLRKNAVSDHV
 VTPPAVKDIKRR

In another embodiment, the bacterial strain used in the process of the invention
 20 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.

25

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.
 30 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.

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
 35 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,

e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:23.

5 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 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.

10

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 to the sequence set forth in SEQ ID NO:27.

15

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 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.

20

25 In another embodiment, the bacterial strain used in the process of the invention 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.

30

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.

35

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 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 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 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 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.

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.

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.

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.

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.

SEQ ID NO: 18:

Nucleotide sequence of ATP synthase gamma chain *atpG_1*

GTGACCGAGCGCCTGTCCGACGTCAACGCCCGCATCGCCTCGGTGCGGCAGCTCTCATCGGT
 20 CATCACGGCCATGCGGGGCATTGCGGCGGCGCGGGCGCGGGAGGCGCGGGGTGCGGCTCGA
 CGGCATCCGCGCCTATGCGCAGACCATCGCCGAGGCCATCGGCCATGTGCTCGCCGTGCTGC
 CCGAGGAGGCCCGCGCCCCGGTCCTCCGGGCACCGGCATCGGGGCCATGCGGTCATCGCCCT
 GTGCGCGGAGCAGGGCTTTGCCGGCGTCTTCAACGAGCGGGTGCTGGACGAGGCCGCCCGG
 CTGCTGACCGGCGGGGCGGGGCCGGCCGAGCTGCTGCTGGTGGGCGACCGGGGCCTGATG
 25 GTGGCCCCGCGAGCGGGGGCTCGATGTCTCCTGGTGGTGGCCATGGTGGCCCATGCGGGCC
 AGGCCTCGGCGCTGGCGGACCGCATCAGCGAGGAGCTCTACCGGCGGATCGATGCGGGACG
 GGTGACGCGGGTGTCTGGTGGTGCACGCCGAGCCCGCCGCGTCCGCCGCCATCGAGACGGTG
 GTGAAAGTGCTGGTGCCGTTGACTTCGCCCCGCTCCCCCTGGCGCGGGTGGCATCCGCCCC
 GCTCATGACCATGCCGCCCGCGGCTGCTGGCCCAGCTGTCGGAGGAATATGTGTTGCCCCG
 30 AGCTGTGCGAGGCGCTCACCTTGTCTTCGCGGCGGAGAACGAGGCCCGCATGCGGGCCAT
 GATCGCCGCCCCGCGCAATGTGGCCGATACCCTGGAGGGCCTCGTCGGCCGCGCCCGGCAG
 ATGCGCCAGGAGGAGATCACCAACGAGATCATCGAGCTGGAAGGCGGCGCCGGCAGCGCCC
 GGCATGCGGATTGA

35 SEQ ID NO: 19:

Amino acid sequence of ATP synthase gamma chain *atpG_1*

MTERLSDVNARIASVRQLSSVITAMRGIAAARAREARGRLDGIRAYAQTIAEAIGHVLAVLPPEAR
ARSSGHRHRGHAVIALCAEQGFAGVFNERNVLDEAARLLTGGAGPAELLVGDRLMVARERGLD
VSWSVPMVAHAGQASALADRISEELYRRIDAGRVTRVSVVHAEPAAASAAIETVVKVLVPFDFARF
PLARVASAPLMTMPPPRLLAQLSEEYVFAELCEALTSFAAENEARMRAMIAARANVADTLEGLVG
5 RARQMRQEEITNEIIELEGGAGSARHAD

SEQ ID NO: 20:

Nucleotide sequence of ATP synthase subunit alpha atpA_1

ATGAGCACGGGCGCGCAAGCGAGCGAGGATTGGCTCACCCGGAGCCGGGCGGCCCTGGCC
10 GGGACGCGCCTTTCCCAGCAATCCCAATCGGTGGGCCGGGTGGAGGAGATGGCCGACGGCA
TCGCCCCGCTCTCCGGCCTGCCGGATGTGCGGCTCGACGAGCTTCTCACCTTCGAGGGCGGC
CAGACCGGCTATGCCCTCACCTCGATCGCACCGAGATCGCCGTGGTGTCTGCTGGATGACGC
CTCCGGCGTGGAGGCGGGCGCCGGGTGTTGCGCACCGGCGAGGTGGTGAAGGTGCCGGT
GGGGCCGGGGCTGCTGGGCCGCATCGTCGACCCCTCGGCCGGCCCATGGACCGCTCCGAG
15 CCGGTGGTGGCGCAGGCGACCATCCCATCGAGCGGCCGGCGCCGGCCATCATCGCCCCGG
ACCTGGTCTCGCAGCCGGTTCAGACCGGCACGCTGGTGGTGGATGCGCTGTTCTCCCTCGGC
CGGGGCCAGCGCGAGCTCATCATCGGCGACCGGGCTACCGGCAAGACCGCCATCGCGGTGG
ACACCATCATCAGCCAGAAGCATTTCGACATCGTGTGCATCTACGTGGCGGTGGGCCAGCGC
GCCGCCGCCGTGGAGCGGGTGGTGGAGGCGGTGCGCGCCACGGGGCGATCGAGCGCTGC
20 ATCTTCGTGGTCGCCTCGGCCCGCCCTCGCCAGGGCTGCAATGGATCGCGCCGTTCCCGG
CATGACCATGGCGGAATATTTCCGCGACAACGGCCAGCATGCGCTCATCATCATGATGATCT
CACCAAGCATGCGGCCACCATCGCGAGCTGGCGCTGCTACCCACGAGCCGCCGGGGCCGC
GAGGCCTATCCCGGCGACATCTTCTATGTGCACGCCCGCCTTCTGGAGCGGGCCGCAAGCT
CTCCGCCGAGCTGGGCGGTGGCTCGCTCACGGCCCTGCCCATCGCGGAGACGGACGCGGGA
25 AACCTCTCCGCCTATATCCCCACCAACCTCATCTCCATACCGATGGGCAGATCGTGTGGAT
TCGCGGCTGTTTCGCGGCCAACCAGCGCCCGGCGGTGGATGTGGGCCTCTCCGTGAGCCGGG
TGGGCGGCAAGGCGCAGCATCCCGCGCTTCGGGCCGTGTCCGGGCGCATCCGGCTCGATTA
TTCCAGTTCCTGGAGCTGGAATGTTACCCGCTTCGGCGGCATACCGATAACCGCGTGAA
GGCGCAGATCACCCGGGGCGAGCGCATCCGCGCGCTGCTACCCAGCCGCGCTTTTCCACCC
30 TGCGCCTTCAGGACGAGGTGGCGCTGCTGGCCGCGCTGGCGGAGGGGGTGTTCGACACTTT
GGCCCCGGGGCTGATGGGCGCCGTGCGTGCCCGCATTCCGGCCAGCTGGATGCGCAGGTG
AAGGACGTGGCCTCGGCCCTCGCCGAGGGCAAGGTGCTGGAGGAGGGCTTGCACGCCCGTC
TCGTGGCGGCCGTGCGGGCCGTGCGGGCGGACGTGGCCGCGACCGCGAAGGCCGGGCCGT
GA

35

SEQ ID NO: 21:

Amino acid sequence of ATP synthase subunit alpha atpA_1

MSTGAQASEDWLTRSRAALAGTRLSQQSQSVGRVEEMADGIARVSGLPDVRDELDTFEGGQT
GYALTDRTEIAVVLLDDASGVEAGARVFGTGEVVKVPVGPGLLGRIVDPLGRPMRSEPVVAQA
HHPIERPAPAIARDLVSQPVTGTLVVDALFSLGRGQRELIIGDRATGKTAIAVDTIISQKHS DIV
CIYVAVGQRAAAVERVVEAVRAHGAIERCIFVVASAAASPGLQWIAPFAGMTMAEYFRDNGQHA
5 LIIIDDLTKHAATHRELALLTHEPPGREAYPGDIFYVHARLLERA AKLSAELGGGSLTALPIAETDAG
NLSAYIPTNLISITDGQIVLDSRLFAANQRPVVDVGLSVSRVGGKAQHPALRAVSGRIRLDYSQFL
ELEMFRFGGITDTRVKAQITRGERIRALLTQPRFSTLRLQDEVALLAALAEGVFDTLAPGLMGAV
RARIPAQLDAQVKDVASALAEKGVLEELHARLVAAVRAVAADVAATAKAGP

10 SEQ ID NO: 22:

Nucleotide sequence of ATP synthase subunit b atpF_1

ATGCAGATCGACTGGTGGACGCTGGGCCTGCAGACGGTCAACGTCCTCGTTCTCATCTGGCT
CCTGAGCCGCTTCCTGTTCAAGCCGGTGGCGCAGGTCATCGCGCAGCGCCGTGCCGAGATCG
AGAAGCTGGTGGAGGATGCGCGCGCCGCCAAGGCCGCCGCCGAGGCCGAGCGGGACACGG
15 CGAAGGCGGAGGAGGCGCGCCTTGCCGCCGAGCGCGCGCCCGCATGGCGGCGGTGCGCA
AGGAGGCGGAGGCGCAGAAGGCGGCATTGCTGGCCGCCGCCAAGACCGAGGCCGAGGGCCC
TGCACGCGGCCGCGGAAGCGGCCATCGTCCGGGCGCGGGCGAGCGAGGAGGAAGCCGCCG
CCGACCGCGCCAGCCGCCTTGCCGTGGACATCGCCGCCAAGCTGCTGGACCGGCTGCCCCA
CGACGCCCCGGTGC GGGCTTCATCGATGGCTCGCCGAGGGGCTTGAAGCCCTGCCCCAG
20 GCGAGCCGGGCGGTGATCGGCGTCGACGGCGCGCCAGTGCGCGTGACGGCCGCGCGCGCC
CTTATGCCGGCGGAGGAGGAGGCCTGCCGCACGCGGCTCTCCAGGCGCTGGGCCGTCCGG
TGACGCTGGCCGTGACCATCGACCCCGCCCTCATCGCCGGCTGGAGATGGAGACGCCCCAC
GCGGTGGTGC GCAATTCCTCAAGGCCGATCTCGACCGCGTCACCGCGGCGCTCACCCATCA
TGGGACCTGA

25

SEQ ID NO: 23:

Amino acid sequence of ATP synthase subunit b atpF_1

MQIDWWTLGLQTVNVLVLIWLLSRFLFKPVAQVIAQRRAEIEKLVEDARA AKA AEAERDTAKAE
EARLAAERGARMAAVAKEAEAQKAALLAAKTEAEALHAAAEAAIVRARASEEEEEAADRASRLAV
30 DIAAKLLDRLPDDARVAGFIDGLAEGLEALPEASRAVIGVDGAPVRVTAARALMPAE EEEACRTRLS
QALGRPVT LAVTIDPALIAGLEMETPHAVVRNSFKADLDRVTAALTHHGT

SEQ ID NO: 24:

Nucleotide sequence of ATP synthase subunit c, sodium ion specific atpE_1

ATGACTGTCGAGATGGTCAGCATCTTCGCGGCGGCGCTCGCCGTCTCCTTCGGCGCCATCGG
35 GCCGGCCCTGGGCGAGGGCCGGGCGGTGGCCGCGGCCATGGACGCCATCGCCCGCCAGCC

GGAGGCGGCCGGAACCTTGTGCGCACGCTCTTCGTCGGCCTCGCCATGATCGAGACCATGG
CGATCTACTGCCTGGTGATCGCGCTCCTGGTGCTCTTCGCCAATCCGTTCTGTAAGTGA

SEQ ID NO: 25:

- 5 Amino acid sequence of ATP synthase subunit c, sodium ion specific atpE_1
MTVEMVSIFAAALAVSFGAIGPALGEGRAVAAAMDIAIRQPEAAGTLSRTLFVGLAMIIETMAIYCL
VIALLVLFANPFVK

SEQ ID NO: 26:

- 10 Nucleotide sequence of ATP synthase subunit a atpB_1
ATGGGCTCGCCGCTGATCCTCGAACCCTGTTCCATATCGGGCCCGTGCCCATCACCGCGCC
GGTGGTGGTCACCTGGCTCATCATGGCCGCCTTCATTGGGCTGGCGCGGCTCATCACCCGGA
AGCTTTCCACCGATCCCACCCGACCCAGGCGGCGGTGGAACGGTGCTGACCGCCATCGAT
TCCCAGATCGCCGACACCATGCAGGCCGATCCCGCGCCTTATCGCGCGCTCATCGGCACCAT
15 CTTCCTTTATGTGCTGGTGGCCAACTGGTCCTCGCTCATCCCGGGCATCGAGCCGCCACGG
CGCATATCGAGACCGATGCGGCGCTCGCTTTCATCGTGTTCCGCCACCATCGGGTTCGGG
TTGAAGACAAGGGGTGTGAAGGGCTATCTCGCCACCTTCGCCGAACCTCCTGGGTGATGAT
CCCGCTCAATGTGGTGGAGCAGATCACCCGACCTTCTCGCTCATCGTGCGCTGTTCGGCA
ACATCATGAGCGGGGTGTTCTGGTGGCATCATCTGTCCCTCGCCGGGCTGCTGGTGCCC
20 ATCCCCCTCATGGCGCTCGATCTCCTGACCGGCGCCGTGCAGGCCTACATCTTCGCGGTGCT
GGCCTGCGTGTTTCATCGGCGCGGCCATTGGCGAGGCGCCGGCAAAGCCCCAATCGAAGGAG
CCAGGGAAAACATCATGA

SEQ ID NO: 27:

- 25 Amino acid sequence of ATP synthase subunit a atpB_1
MGSPLILEPLFHIGVPITAPVVVTLIMAAFIGLARLITRKLSTDPTRTQAAVETVLTADSQIADT
MQADPAPYRALIGTIFLYVLVANWSSLIPGIEPPTAHIEDAALAFIVFAATIGFGLKTRGVKGYLAT
FAEPSWVMIPLNVVEQITRTFSLIVRLFGNIMSGVFVVGIIISLAGLLVPIPLMALDLLTGAVQAYIF
AVLACVFIGAAIGEAPAKPQSKEPGKTS

30

SEQ ID NO: 28:

- Nucleotide sequence of ATP synthase epsilon chain atpC_1
GTGAGCGCGCCGCTGCACCTCACCATCACACGCCGCGCCGTTCTGGTGGACCGTGCCGA
CATCGTGGCCCTGCGTGCCGAGGACGAGAGCGGCAGCTTCGGCATCCTGCCCGGCCATGCG
35 GATTTCTGACCGTTCTGGAGGCCTGCGTGGTGCCTTCAAGGATGGGGCCGACGGCGTGCA
TTATTGTGCTCTCAGTGGTGGCGTGCTGTGCGTGCAGGAGGGCCGGCGCATCGCCATCGCCT
GCCGTCAGGGCACGGTGAGCGACGACCTGGTCGCCCTGGAAGGGGCGGTGGACGCCATGC

GTTCGGCGGAGAGCGATGCCGACAAGCGGGCCCGGGTGGAGCAGATGCGCCTTCATGCCCA
CGCCGTGCGCCAGCTCCTGCACTATCTGCGGCCCGGCCGGCGGCGGCGTGGCGCCGGCC
GCCGCGCCGGAGGAGGGGCCGTCATGA

5 SEQ ID NO: 29:

Amino acid sequence of ATP synthase epsilon chain atpC_1

MSAPLHLTITTPAAVLVDRADIVALRAEDESFSFILPGHADFLTVLEACVVRFKDGADGVHYCAL
SGGVLSVEEGRRIAIACRQGTVSDDLVALEGAVDAMRSAESDADKRARVEQMRLHAHAVRQLL
HYLRPGRAGGVAPAAPEEGPS

10

SEQ ID NO: 30:

Nucleotide sequence of ATP synthase subunit beta atpD_1

ATGGCAGCGGCAGATGAGGAGGCGCAATCGGCCCGGCCCGCCTCGGGCCGGGTGGTG
GCCGTGCGCGGCGCGGTGATCGACATCGCCTTTGCCAGCCTCCGCTGCCGCCGCTGGACG
15 ACGCCCTTCTCATCACCGACGGCCGGGGCGGCACGGTGCTGGTGGAGGTGCAGAGCCATAT
GGATCGGCACACGGTGCGCGCCATCGCCCTTCAGGCCACCACCGGCCTCAGCCGGGGGCTG
GAGGCGGCGCGGGTGGGCGGGCCGGTGAAGGTGCCGGTGGGAGACCATGTGCTCGGCCGC
CTCCTGGATGTACCGGGCGCCATCGGCGACAAGGGCGGGCCGCTGCCGCCGACGTGCCCA
CGCGGCCGATCCACCACGCGCCGCCATCCTTCGCCGCGCAGGGCGGCACGTCCGATCTGTTT
20 CGCACCGGCATCAAGGTCATCGACCTCCTGGCGCCCCTCGCCCAGGGCGGCAAGGCGGCCA
TGTTGCGCGGGGCCGGCGTGGGCAAGACCGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GGCGAGCTACAAGGGCATCTCGGTGTTTGCCGGCGTGGGGGAGCGCTCCCGCGAGGGCCAC
GAGATGCTGCTGGACATGACCGATTCCGGCGTGCTCGACCGCACCGTTCTGGTCTATGGCCA
GATGAACGAGCCCCCGGGGCCCGCTGGCGGGTGGCCATGACGGCGCTGACCATCGCCGAA
25 TATTTCGCGACGAGAAGCACCAGAAGCTCCTGCTGCTGATGGACAACATCTTCCGCTTCGTC
CAGGCGGGGGCGGAGGTCTCCGGCCTTTTGGGCCGTCCGCCCTCCCGGGTGGGATACCAGC
CGACGCTGGCGAGCGAGGTGGCGGCGCTCCAGGAACGCATCACCTCCGTGGGCGAGGCCTC
GGTGACCGCCATCGAGGCGGTCTACGTGCCGGCGGATGACTTCACCGATCCCGCCGTGACCA
CCATCGCCGCCACGTGGATTCCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
30 GTATCCGGCGGTGGACCCCATCTCCTCCTCGTCGGTGCTGCTCGACCCGCTCATCGTGGGGG
ACGAGCATGCGCGCGTCGCCAACGAGGTGCGCCGGACCATCGAGCATTATCGCGAGCTTCAG
GATGTGATCTCGCTGCTGGGCATGGAGGAATTGGGCACCGAGGATCGCCGCATCGTGGAGC
GGGCGCGCCGGCTCCAGCGCTTCCTACCCAGCCCTTCACGGTCACCGAGGCCTTCACCGGC
GTGCCCGGCCGCTCGGTGGCCATCGCCGACACCATCGCCGGCTGCAGGATGATCCTGTCCG
35 GCGCCTGCGACGACTGGCAGGAAAGCGCCCTCTACATGGTGGGCACCATCGACGAGGCCCCG
CCAGAAGGAGGAGGCCGCTCGCGCCAAGGCGGGGCGAGGGCGCCCCGGCGGGACGGCAGC
CGAGACGGCGGAGGCCGCCCGGTGA

SEQ ID NO: 31:

Amino acid sequence of ATP synthase subunit beta atpD_1

MAAADEEAQSAAGPASGRVVAVRGAVIDIAFAQPPLPLDDALLITDGRGGTVLVEVQSHMDRH
5 TVRAIALQATTGLSRGLEAARVGGPVKVPVGDHVLGRLLDVTGAIGDKGGPLPADVPTRPIHHAP
PSFAAQGGTSDLFRTGIKVIDLLAPLAQGGKAAMFGGAGVGKTVLMELIHAMVASYKGISVFAG
VGRSREGHEMLLDMTDSGVLDRTVLVYGQMNEPPGARWRVPMTALTIAEYFRDEKHQNVLLLM
DNIFRFVQAGAEVSGLLGRPPSRVGYQPTLASEVAALQERITSVGEASVTAIEAVYVPADDFTDPA
VTTIAAHVDSMVLSRAMAAEGMYPVDPISSSSVLLDPLIVGDEHARVANEVRRRTIEHYRELQD
10 VISLLGMEELGTEDRRIVERARRLQRFLTQPFTVTEAFTGVPGRSVAIADTIAGCRMILSGACDDW
QESALYMVGTIDEARQKEEAARAKAGQGAPAGTAAETAEAAP

SEQ ID NO: 32:

Nucleotide sequence of ATP synthase subunit beta atpD_2

15 ATGGCGAACAAGGTCGGACGCATCACCCAGATCATCGGCGCCGTCGTCGACGTGCAGTTCGA
CGGGCATCTGCCGGCGATTCTCAACGCGATCGAGACCACCAACCAGGGCAACCGGCTGGTG
TCGAAGTGGCTCAGCATCTCGGCGAGAACACCGTGCGCTGCATCGCCATGGATGCCACTGAA
GGCCTGGTGCGTGGCCAGGAGGTGGCCGACACCGATGCGCCCATCCAGGTGCCCCGTGGGCG
CCGCCACCTCGGCCGCATCATGAACGTGATCGGCGAGCCGGTGGACGAGCTGGGCCCCAT
20 CGAGGGCGAAGCGCTGCGCGGCATCCATCAGCCGGCCCCCTCCTATGCGGAGCAGGCCACG
GAAGCTGAGATCCTCGTCACCGGCATCAAGGTGGTGGATCTGCTGGCGCCCTATTCCAAGGG
CGGCAAGGTGGGCCTGTTGCGCGGCGCCGGCGTGGGCAAGACCGTGCTCATCATGGAGCTG
ATCAACAACGTGGCCAAGGCGCACGGCGGCTATTCCGTGTTGCGCGCGTGGGTGAGCGCA
CCC GCGAGGGCAACGACCTCTACCACGAGATGATCGAGTCCAACGTGAACAAGGACCCGCAC
25 GAGAACAATGGCTCGGCGGCCGTTCCAAGTGCGCCCTGGTCTATGGCCAGATGAACGAGCC
GCCCCGGCGCCCGCGCCCGCTGGCCCTCACCGGCCTCACCGTCGCCGAGCATTTCCGCGAC
CAGGGCCAGGACGTGCTGTTCTTCGTGGACAACATCTTCCGCTTCACCCAGGCGGGCTCCGA
GGTGTGCGCGCTTCTCGGCCGCATCCCCTCGGCGGTGGGCTACCAGCCGACGCTGGCCACC
GACATGGGCCAGCTGCAGGAGCGCATCACCAACCACCAAGGGCTCCATCACCTCGGTGCA
30 GGCCATCTACGTGCCGGCGGACGATCTGACCGATCCGGCGCCGGCCGCTCCTTCGCCCATC
TGGACGCCACCACGGTGTGTCGCGCTCCATCGCGGAGAAGGGCATCTACCCGGCGGTGGA
TCCGCTGGACTCCACCTCGCGCATGCTGTCTCCCGCCATCCTCGGCGACGAGCACTACAACAC
CGCGCGCCAGGTGCAGCAGACCCTGCAGCGCTACAAGGCGCTCCAGGACATCATCGCCATCC
TGGGCATGGACGAACCTCTCCGAAGAGGACAAGCTCACCGTGCGCCGCGCCCGCAAGATCGA
35 GCGCTTCCTCTCCAGCCCTTCCACGTGGCCGAGGTGTTACCGGTTGCCCCGGCAAGCTGG
TCGACCTCGCCGACACCATCAAGGGCTTCAAGGGCTGGTGGACGGCAAGTACGACTACCTG

CCCGAGCAGGCCTTCTACATGGTGGGCACCATCGAAGAAGCCATCGAGAAGGGCAAGAAGCT
GGCGGCCGAGGCGGCCTGA

SEQ ID NO: 33:

- 5 Amino acid sequence of ATP synthase subunit beta atpD_2
MANKVGRITQIIGAVVDVQFDGHLPAILNAIETTNGNRLVLEVAQHLGENTVRCIAMDATEGLV
RGQEVADTDAPIQVPVGAATLGRIMNVIGEPVDELGPiEGEALRGIHQPAPSYAEQATEAEILVTG
IKVVDLLAPYSKGGKVGLFGGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERREGNDLYHEMI
ESNVNKDPHENNGSAAGSKCALVYGQMNEPPGARARVALTGLTVAEHFRDQGQDVLFFVDNIFR
10 FTQAGSEVSALLGRIPSAVGYPQLATDMGQLQERITTTTKGSITSVQAIYVPADDLTDPAASFA
AHLDTTVLSRSIAEKGIYPAVDPLDSTRMLSPAILGDEHYNTARQVQQLQRYKALQDIILG
MDELSEEDKLTVARARKIERFLSQPFHVAEFTGSPGKLVLDADTIKGFGLVDGKYDYLPEQAFY
MVGTIIEEAIEKGKKLAEEAA

15 SEQ ID NO: 34:

Nucleotide sequence of ATP synthase gamma chain atpG_2

- ATGGCGAGTCTGAAGGACCTGAGAAACCGCATTGCCTCGGTGAAGGCGACGCAGAAGATCAC
CAAGGCGATGCAGATGGTCGCCGCGGCGAAGCTGCGTCGCGCCAGGCGGCGGCTGAAGC
GGCCCGTCCCTATGCGGAACGCATGGAGACGGTGCTCGGAAATCTTGCCTCCGGCATGGTGG
20 TGGGCGCGCAGGCGCCTGTTCTCATGACCGGGACGGGCAAGAGCGACACCCACCTGCTGCT
GGTGTGCACCGGCGAGCGCGCCTGTGCGGCGCCTTCAACTCGTCCATCGTGCGCTTCGCCC
GCGAGCGGGCGCAGCTGCTGCTGGCCGAGGGCAAGAAGGTGAAAATCCTGTGCGTGGGCCG
CAAGGGCCACGAGCAGCTGCGCCGCATCTACCCGGACAACATCATCGACGTGGTGGACCTGC
GCGCGGTGCGCAACATCGGCTTCAAGGAGGCCGACGCCATCGCCCGCAAGGTGCTGGCCCT
25 GCTCGATGAAGGCGCATTGACGTCTGCACGCTCTTCTACTCCCACTTCAGGAGCGTGATCGC
CCAGGTGCCGACGGCCAGCAGCTCATTCCGGCCACCTTCGACGAGCGGCCGCGCGTCCGCC
GATGCGCCGGTCTATGAATATGAGCCGGAGGAGGAGATCCTCGCCGAGCTGCTGCCGC
GCAACGTGGCGGTGACATCTTCAAGGCCCTCCTCGAGAACCAGGCTTCTTCTATGGCTCCC
AGATGAGCGCCATGGACAACGCCACGCGCAATGCGGGCGAGATGATCAAGAAGCAGACGCT
30 CACCTACAACCGTACCCGCCAGGCCATGATCACGAAGGAACATCGAGATCATCTCCGGCG
CCGAGGCCGTCTGA

SEQ ID NO: 35:

Amino acid sequence of ATP synthase gamma chain atpG_2

- 35 MASLKDLNRNRIASVKATQKITKAMQMVAALKLRRQAQAAEAARPYAERMETVLGNLASGMVVGA
QAPVLMGTGKSDTHLLLVCTGERGLCGAFNSSIVRFARERAQLLAEGKKVKILCVGRKGHEQL
RRIYPDNIIDVVDLRAVRNIGFKEADAIARKVLALLDEGAFDVCTLFYSHFRSVIAQVPTAQQLIPA

TFDERPAVADAPVYEYEPEEEEEILAEELLPRNVAVQIFKALLENQASFYGSQMSAMDNATRNAGEMI
KKQTLTYNRTRQAMITKELIEIISGAEEV

SEQ ID NO: 36:

- 5 Nucleotide sequence of ATP synthase subunit alpha atpA_2
ATGGACATTCGAGCCGCTGAAATCTCTGCCATCCTGAAAGAGCAGATCCAGAATTTGGGCCAG
GAGGCGGAAGTCTCCGAGGTGGGTGAGTTCTGTCCGTGGGTGACGGCATCGCGCGCGTCT
ACGGCCTCGACAACGTCCAGGCGGGCGAGATGGTCGAGTTCGAGAACGGCACGCGCGGCAT
GGCGCTGAACCTCGAGCTCGACAATGTGGCATCGTGATCTTCGGTTCGACCGCGAGATCA
10 AGGAAGGCCAGACCGTCAAGCGGACCGGCGCCATCGTGACGCCCCCGTCGGCAAGGGCCT
GCTCGGCCGCGTCGTGGACGCTCTCGGCAACCCGATCGACGGCAAGGGCCCGATCATGTTCA
CCGAGCGTCGCCGGGTGACGTGAAGGCGCCGGGCATCATCCCGCGCAAGTCGGTGCACGA
GCCCATGCAGACCGGCCTGAAGGCCATCGATGCGCTCATCCCCATCGGCCGCGGCCAGCGC
GAGCTCATCATCGGCGACCGCCAGACCGGCAAGACCGCCGTGGCGCTCGACTCGATCCTGAA
15 CCAGAAGCCCATCAACCAGGGCGACGACGAGAAGGCCAAGCTCTACTGCGTCTATGTCGCGG
TGGGCCAGAAGCGTTCCTACTGTCGCGCAGTTCTGTAAGGTGCTCGAGGAGCACGGCGCGCT
GGAATATTCCATCGTCGTCGCCGCCACCGCCTCGGACGCGGCCCCCATGCAGTTCCTGGCGC
CGTTCACCGGCACCGCCATGGGCGAGTATTTCCGCGACAACGGCATGCACGCCCTCATCATC
CATGATGACCTGTCCAAGCAGGCCGTGGCCTACCGCCAGATGTCGCTGCTGCTGCGCCGCCC
20 GCCGGGCCGCGAGGCCTATCCCGGCGATGTGTTCTACCTGCACTCCCGCCTCTTGAGCGCG
CCGCCAAGCTCAATGACGAGCACGGCGCCGGCTCGCTGACCGCCCTGCCGGTGATCGAGAC
CCAGGCCAACGACGTGTCGGCCTACATCCCGACCAACGTGATCTCCATCACCGACGGTCAGA
TCTTCCTTGAATCCGATCTGTTCTACCAGGGCATCCGCCCCGGCGGTGAACGTGGGCCTGTG
GTGTCGCGCGTGGGCTCTTCGGCCAGATCAAGGCGATGAAGCAGGTGGCCGGAAGATCA
25 AGGGCGAGCTCGCCAGTATCGCGAGCTGGCGGCCTTCGCCAGTTCGGTTGCGACCTGGA
CGCGGCCACCCAGAAGCTGCTGAACCGCGGCGCCCGCCTCACCGAGCTGCTGAAGCAGAGC
CAGTTCTCGCCCCTCAAGGTGGAGGAGCAGGTGGCGGTGATCTATGCCGGCACCAATGGCTA
TCTCGATCCGCTGCCGGTCTCCAAGGTGCGCGAGTTCGAGCAGGGTCTGCTCCTGTGCTGC
GCTCGCAGCATCCGGAGATCCTGGACGCCATCCGCACGTCCAAGGAGCTTTCCAAGGACACC
30 GCCGAGAAGCTGACGAAGGCCATCGACGCCTTCGCCAAGAGCTTCTCCTGA

SEQ ID NO: 37:

- Amino acid sequence of ATP synthase subunit alpha atpA_2
MDIRAAEISAILKEQIQNFGQEAHVSEVGVLSVGDGIARVYGLDNVQAGEMVEFENGTRGMAL
35 NLELDNVGIVIFGSDREIKEGQTVKRTGAIVDAPVGKLLGRVVDALGNPIDGKGPIMFTERRRV
DVKAPGIIPRKSVEHPMQTGLKAIDALIPIGRGQRELIIGDRQTGKTAVALDSILNQKPINQGDDE
KAKLYCVYVAVGQKRSTVAQFVKVLEEHGALEYISIVVAATASDAAPMQFLAPFTGTAMGEYFRDN

GMHALIIHDDLSKQAVAYRQMSLLRRPPGREAYPGDVFYLSRLLERA AKLNDEHGAGSLTALP
VIETQANDVSAYIPTNVISITDGQIFLESDFYQGIRPAVNVGLSVSRVGSSAQIKAMKQVAGKIK
GELAQYRELA AFAQFGSD LDAATQKLLNRGARL TELLKQSQFSPLKVEEQVAVIYAGTNGYLDPLP
VSKVREFEQGLLLSLRSQHPEILDAIRTSKELSKDTAEKLTAKIDAFKSF5

5

SEQ ID NO: 38:

Nucleotide sequence of ATP synthase subunit delta atpH

GTGGCGGAAACGATCGTGT CAGGCATGGCGGGACGCTATGCGACCGCGCTGTT CAGCTGG
CGGACGAAGCCGGTGCCATCGATTCCGTCCAGGCGGATCTTGATCGCCTGTCCGGCCTTCTG
10 GCCGAGAGCGCGGATCTGGCGCGGCTGGTCAAGAGCCCGGTCTTCACCGCCGAGCAGCAGC
TCGGCGCGATGGCGGCCATTCTCGATCAAGCAGGCATTTCCGGCCTTGCGGGCAAATTCGTG
AAGCTGGTGGCGCAGAACCGCCGCCTGTT CGCACTGCCGCGCATGATTGCCGAATACGCCGT
CCTGGTGGCCCGGAAGAAGGGCGAGACCTCGGCGAGCGTGACCGTTGCCACCCCCCTGAGC
GATGAGCATCTGGCCACGCTCAAGGCGGCCCTGGCTGAAAAGACCGGCAAGGACGTGAAGC
15 TCGACGTCACCGTCGATCCGTCCATCCTCGGTGGTCTCATCGTGAAGCTCGGCTCGCGCATG
GTCGATGCTTCCCTGAAGACCAA AACTCAATTCTATCCGGCATGCGATGAAAGAGGTCCGCTGA

SEQ ID NO: 39:

Amino acid sequence of ATP synthase subunit delta atpH

MAETIVSGMAGRYATALFELADEAG AIDSVQADLDRLSGLLAESADLARLVKSPVFTAEQQLGAM
20 AAILDQAGISGLAGKFVKLV AQNRRLFALPRMIAEYAVLVARKKGETSASVT VATPLSDEHLATLK
AALAEKTGKDVKLDVTVDPSILGGLIVKLGSRMVDASLTKLNSIRHAMKEVR

SEQ ID NO: 40:

25 Nucleotide sequence of ATP synthase subunit b atpF_2

ATGACCGAAATGGA ACTGGCTGAGCTCTGGGTCGCCATCGCCTTCCTGGTTTTCTG TAGGCCTC
CTGATCTATGCGGGCGCCACCGCGCCATCGTCTCCGCCCTGGATTCCCGCGGCTCGCGCAT
CGCCTCGGAACTGGAGGAGGCCCGTCGGCTCAAGGAAGAGGCC CAGAAGCTGGTGGCCGAA
TTCAAGCGCAAGCAGCGCGAGGCCGAGGCCGAGGCCGAATCCATCGTCACCGGCGCCAAGG
30 CCGAGGCCGAGCGCCTCGCCGCCGAGGCCAAGGCGAAGATCGAGGATTTCTGTCACCCGCCG
CACCAAGATGGCCGAGGACAAGATCGCCAGGCCGAGCATCAGGCTCTGGCGGACGTGAAG
TCCATCGCCGCCGAGGCGGCGGCCAAGGCGGCCGAGGTGATCCTCGGCGCCCAGGCCACCG
GCGCGGTGGCGGAGCGTCTGCTGTCTGGGCGCCATCTCCGAGGTCAAGACCAAGCTCAACTG
A

35

SEQ ID NO: 41:

Amino acid sequence of ATP synthase subunit b atpF_2

MTEMELAEELWVAIAFLVFVGLLIYAGAHRAIVSALDSRGSRIASELEEARRLKEEAQKLVAEFKRK
QREAEAEAESIVTGAKAEAERLAAEAKAKIEDFVTRRTKMAEDKIAQAEHQALADVKSIAAEAAA
KAAEVILGAQATGAVAERLLSGAISEVKTCLN

5 SEQ ID NO: 42:

Nucleotide sequence of ATP synthase subunit b' atpG_3

ATGATGATTGCATGGAAGCGGACCTTCGCAGTCGTGACCTTCGGGGCCGCCCTGATGGCCAT
GCCCCGTCGCGGGCGTGGTCGCAGCTGAGACTTCTCCCGCTCCGGCGGCAGTGGCGCAGGCC
GATCATGCGGTGCCCCACCGAGGCGGCCGCGCCAGGGCACCGCCGATGCGGCCCATGCCGCCG
10 CGCCGGGCGAGGCCGCCCATGGTGGCGCGGCCAAGCACGAAACCCATTTCCCGCCCTTCGA
CGGCACCACCTTCGCCTCCAGTTGCTGTGGCTCGCCGTCACCTTCGGCCTGCTTTACTACCT
CATGAGCAAGGTCACGCTGCCGCGCATCGGCCGCATCCTGGAAGAGCGCCACGACCGCATC
GCCGATGATCTGGAGGAAGCCTCCAAGCATCGCGCCGAGAGCGAGGCCGCCAGCGGGCCT
ATGAGAAGGCGCTGAGCGAGGCCCGCGCGAAGGCCCATTCATCGCCGCGGAAACCCGCGA
15 CCGCCTTGCCGCCACGCCGACACCAACCGCAAGGCGCTGGAGAGCGAGCTCACCGCCAAG
CTGCAGGCGGCCGAGGAGCGCATCGCCACCACCAAGAGCGAAGCCCTCACCCATGTGCGCG
GCATCGCGGTGGACGCCACCCAATCCATCGTCTCCACCCTCATCGGTGTGCGCCCCGCGGCG
GCCGACGTGAAAAAGCGGTGGACGCGCCCTGTCCCAGCACGCCAGGCCTGA

20 SEQ ID NO: 43:

Amino acid sequence of ATP synthase subunit b' atpG_3

MMIAWKRTFAVVTFGAALMAMPVAGVVAETSPAPAAVAQADHAVPTEAAGQGTADAHAHAAP
GEAAHGGAAKHETHFPPFDGTTFASQLLWLA VTFGLLYLMSKVTLPRIGRILEERHDRIADDDLEE
ASKHRAESEAAQRAYEKALSEARAKAHSIAAETDRDLAAHADTNRKALESELTKLQAAEERIATT
25 KSEALTHVRGIAVDATQSIVSTLIGVAPAAADVEKA VDGALSQHGQA

SEQ ID NO: 44:

Nucleotide sequence of ATP synthase subunit c atpE_2

ATGGAAGCGGAAGCTGGAAAGTTCATCGGTGCCGGCCTCGCCTGCCTCGGCATGGGTCTCGC
30 TGGCGTCGGCGTCGTAACATCTTCGTAACCTCTCTCCGGCGCCCTGCGCAACCCGTCGG
CTGCCGACGGCCAGTTCGCCCCGCGCCTTCATCGGCGCCGCCCTGCGGAAGGTCTCGGCATC
TTCTCGCTGGTCGTTGCGCTCGTCCTGCTGTTCTGTTGGCCTGA

SEQ ID NO: 45:

35 Amino acid sequence of ATP synthase subunit c atpE_2

MEAEAGKFIGAGLACLGMLAGVGVGNIFGNFLSGALRNPSAADGQFARAFIGAALAEGLGIFSL
VVALVLLFVA

SEQ ID NO: 46:

Nucleotide sequence of ATP synthase subunit a atpB_2

ATGACCGTCGATCCGATCCACCAGTTCGAGATCAAGCGCTACGTGGATCTGCTGAACGTCGG
5 CGGTGTCCAGTTCTCCTTACCAACGCAACGGTGTTTCATGATTGGCATCGTCCTGGTGATTTT
CTTCTTCTGACTTTCGCGACACGCGGTGCGACCCCTTGTCGGGGCCGGATGCAGTCGGCGG
CGGAGCTGAGCTACGAGTTCATCGCCAAGATGGTGCGCGACGCGGCCGGCAGCGAGGGAAT
GGTGTTCTTTCCCTTCGTCTTCTCGCTCTTCATGTTTCGTGCTGGTGGCGAACGTATTGGGGCT
CATCCCCTACACCTTCACGGTGACCGCCCCACCTCATCGTCACCGCCGCCCTGGCGGCGACGG
10 TGATCCTCACCGTCATCATCTACGGCTTCGTGCGGCACGGCACCCACTTCCTGCACCTGTTTCG
TGCCGTGCGGCGTGCCGGGCTTCTCCTGCCCTTCTCGTGGTGATCGAGGTGGTGTCGTTTC
CTGTGCGGCCCCATCAGCCTCTCGCTGCGTCTGTTTCGCCAACATGCTGGCGGGCCACATCGC
CCTCAAGGTGTTGCGCTTCTTCGTGCTGGGACTGGCCTCGGCCGGCGCGATCGGCTGGTTTCG
GCGCCACCCTGCCCTTCTTCATGATCGTGCGCTCACCGCGCTGGAGCTGCTGGTGGCGGTG
15 CTGCAGGCCTACGTGTTGCGGGTGCTGACCTCGATCTACCTCAACGACGCCATCCATCCCCGGC
CACTGA

SEQ ID NO: 47:

Amino acid sequence of ATP synthase subunit a atpB_2

MTVDPIHQFEIKRYVDLLNVGGVQFSFTNATVFMIGIVLVIFFLTFATRGRTLVPGRMQSAAELSY
20 EFIAKMVRDAAGSEGMVFFPFVFSLFMFVLVANVLGLIPYTFVT AHLIVTAALAATVILTVIIYGFV
RHGTHFLHLFVPSGVPGLLPFLVVEVVSFLSRPISLSLRLFANMLAGHIALKVFAFFVVG LASAGA
IGWFGATLPFFMIVALTALELLVAVLQAYVFAVLTSIYLNDAIHPGH

25 SEQ ID NO: 48:

Nucleotide sequence of ATP synthase protein I atpI

ATGTCCGAGCCGAATGATCCATCCCGCAGGGACGGTGCGAAGGCGAAAGACGAGACGCAGG
ACTCCCGGCCCCGGTGAGGCGGATCTTGCTCGGCGCCTCGATGCGCTCGGCACCTCCATCGGT
CAGGTCAAGTCCAGAAGCGGGGAGCCCGCGGCGACGCCGCGCAAGGACACCTCCTCGGCCT
30 CCGGCGCGGCCCTGGCGTTTCGGCTGGGCGCCGAGTTTGTTCAGGCGTGCTGGTGGGCTC
GCTCATCGGCTACGGGTTGGATTATGCGTTTGCGATTCGCCCTGGGGGCTGATCGCCTTCAC
GCTGATCGGCTTTGCCGCCGGCGTCCTGAACATGCTGCGCGTGGCGAACAGCGATGCCAAGC
GCCACAGCGCGGACAGGTGA

35 SEQ ID NO: 49:

Amino acid sequence of ATP synthase protein I atpI

MSEPNDPSRRDGAKAKDETQDSRPGEADLARRLDALGTSIGQVKSRSGEPAATPRKDTSSASG
AALAFRLGAIEFVSGVLVGSLIGYGLDYAFAISPWGIIAFTLIGFAAGVLNMLRVANSDAKRHSADR

In another embodiment, the bacterial strain used in the process of the invention
5 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.
more than 98%, such as more than 99% sequence identity to the sequence set forth in
10 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
having the sequence set forth in SEQ ID NO:53 or a sequence having more than 60%
15 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
comprises a gene encoding a nitrogenase molybdenum-iron protein beta chain nifK_1
20 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.

25 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
98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID
30 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
35 sequence set forth in SEQ ID NO:59.

SEQ ID NO: 50:

Nucleotide sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_1

ATGAGTTCGCTCTCCGCCACTATTCAACAGGTCTTCAACGAGCCGGGCTGCGCGAAGAACCA
GAATAAGTCCGAGGCGGAGAAGAAGAAGGGCTGCACCAAGCAGCTGCAACCCGGCGGAGCG
GCCGGCGGCTGCGCGTTTCGACGGCGCGAAGATCGCGCTCCAGCCCTTGACCGACGTGCGCC
5 ACCTGGTGCACGGCCCCATCGCCTGCGAAGGCAATTCTGGGACAATCGTGGCGCCAAGTCC
TCCGGCTCGAACATCTGGCGCACCGGCTTACCACGGACATCAACGAAACCGACGTGGTGT
CGGCGGCGAGAAGCGTCTGTTCAAGTCCATCAAGGAAATCATCGAGAAGTACGACCCGCCGG
CCGTCTTCGTCTATCAGACCTGCGTCCCCGCCATGATCGGCGACGACATCGACGCGGTGTGC
AAGGCGGCCAGGGAGAAGTTCGGAAAGCCGGTGATCCCGATCAATTCCCCCGGCTTCGTGG
10 GGCCGAAGAATCTCGGCAACAAGCTCGCCGGCGAGGCGCTCCTCGACCATGTGATCGGCACC
GAGGAGCCCGATTACACGACGGCCTACGACATCAACATCATCGGCGAATACAATCTCTCCGG
CGAGTTGTGGCAGGTGAAGCCGCTGCTGGACGAGCTGGGCATCCGCATCCTCGCTGCATCT
CCGGCGACGGGAAGTACAAGGATGTGGCGTCTCCACCGCGCCAAGGCGGCGATGATGGT
GTGCTCCAAGGCCATGATCAACGTGGCCCGCAAGATGGAGGAGCGCTACGACATCCCCTTCT
15 TCGAAGGCTCCTTCTACGGCATCGAGGATAGCTCCGATTCCCTGCGCGAGATTGCGCGCATG
CTCATCGAGAAGGGCGCCGATCCGGAGCTGATGGACCGCACCGAGGCGCTGATTGAGCGGG
AAGAGAAGAAGGCGTGGGACGCCATCGCCGCCTACAAGCCCCGCTTCAAGGACAAGAAGGT
GCTGCTCATCACCGGCGGCGTGAAATCCTGGTGGTGGCAGCGCTCCAGGAAGCCGGC
CTCGAACTGGTGGGCACCTCGGTGAAGAAGTCCACCAAGGAGGACAAGGAGCGCATCAAGG
20 AACTGATGGGCCAGGACGCCCACATGATCGACGACATGACGCCCCGCGAAATGTACAAGATG
CTGAAGGACGCCAAGGCGGACATCATGCTCTCGGGCGGGCGCTCGCAATTCATCGCGCTCAA
GGCCGCCATGCCCTGGCTCGACATCAACCAGGAGCGCCACCACGCCTATATGGGCTATGTGG
GCATGGTGAAGCTGGTCGAGGAGATCGACAAGGCGCTCTACAATCCCGTGTGGGAACAGGT
GCGCAAGCCCCGCCCGTGGGAAAATCCGGAAGACACCTGGCAGGCCCCGTGCGCTCGCCGAA
25 ATGGAGGCGGAGGCCGCCGCTCGCCGCCGATCCGGTGCAGCGCGGAAGAGGTGCGCCGG
TCCAAGAAGATCTGCAATTGCAAGAGCGTCGACCTCGGAACCATGAGGACGCCATCAAGGC
TCACGCGCTGACCACCGTGGAGGGTGTGCGAGAGCACACCAATGCCTCGGGAGGCTGCGGA
GCCTGCAGCGGGCGGATCGAGGAGATCTTCGAGGCCGTGGGCGTTGTCGCCGCCCGCCTC
CCGCGGAGGCCGCCCGTCTCCGCAGGAGATCGCGCCCGATCCGCTCGCTGCGGAGGAAAA
30 GCGCCGCGCCAAGAAGGCCTGCGGCTGCAAGGAGGTAGCGGTGCGCACCATGAGGATGCC
ATCCGCGCCAAGGGTCTGCGAAACATCGCGGAGGTGCGTGCGGCCACCGATGCCAACCCG
GCTGCGGCAATTGCCAGGAGCGGGTGGAGGGCATCCTCGACCGGGTTCTCGCCGAGGCGGC
CTCAGAACTCCAGGCGGCGGAATAG

35 SEQ ID NO: 51:

Amino acid sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_1

MSSLSATIQQVFNEPGCAKNQNKSEAEKKKGCTKQLQPGGAAGGCAFDGAKIALQPLTDVAHLV
HGPIACEGNSWDNRGAKSSGSNIWRTGFTTDINETDVVFGGEKRLFKSIKEIEKYDPPAVFVYQ
TCVPAMIGDDIDAVCKAAREKFGKPVIPINSPGFVGPKNLGNKLAGEALLDHVIGTEEPDYTTAYD
INIIGEYNLSGELWQVKPLDELGIRILACISGDGKYKDVASSHRAKAAMMVCSKAMINVARKME
5 ERYDIPFFEGSFYGIEDSSDSLREIARMLIEKGADPELMDRTEALIEREEKAWDAIAAYKPRFKDK
KVLITGGVKSWSVVAALQEAGLELVGTSVKKSTKEDKERIKELMGQDAHMIDDMTPREMYKML
KDAKADIMLSGGRSQFIALKAAMPWLDINQERHHAYMGYVGMVKLVEEIDKALYNPVWEQVRKP
APWENPEDTWQARALAEMEAEEAALAADPVRAEEVRRSKKICNCKSVDLGTIEDAIKAHALTTVE
GVREHTNASGGCGACSGRIEEIFEAVGVVAAPPPAEAAPSPQEIAPDPLAAEEKRRRAKKACGCKE
10 VAVGTIEDAIRAKGLRNIAEVRAATDANTGCGNCQERVEGILDRVLAEAASELQAAE

SEQ ID NO: 52:

Nucleotide sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_2

ATGAGTGTGCGACAGTCCCAGAGCGTCGCCGAGATCAAGGCGCGCAACAAGGAACATCATCGA
15 AGAGGTCTCAAGGTCTATCCCGAGAAGACCGCCAAGCGCCGCGCCAAGCACCTGAACGTCC
ACGAAGCCGGCAAGTCCGACTGCGGCGTGAAAGTCCAACATCAAGTCCATCCCGGGCGTGATG
ACCATCCGCGGTTGCGCTTATGCCGGCTCCAAGGGTGTGGTGTGGGGTCCCATCAAGGACAT
GATCCACATCTCCACGGCCCCGTGGGCTGCGGCCAGTATAGCTGGGCCGCCCGCCGCAACT
ACTATATCGGCACGACCGGCATCGACACCTTCGTGACGATGCAGTTCACCTCCGACTTCCAGG
20 AGAAGGACATCGTCTTCGGCGGCGACAAGAAGCTCGCCAAGATCATGGACGAGATCCAGGAG
CTGTTCCCGCTGAACAACGGCATCACCGTTCAGTCCGAGTGCCCCATCGGCCTCATCGGCGA
CGACATCGAGGCCGTCTCCAAGCAGAAGTCCAAGGAGTATGAGGGCAAGACCATCGTGCCGG
TGCGCTGCGAGGGCTTCGCGGCGGTGTCAGTCCCTGGGCCACCACATCGCCAACGACGCC
ATCCGCGATTGGGTGTTGACAAGATCGCGCCGACGCCGAGCCGCGCTTTGAGCCGACCCC
25 GTACGACGTGCGCATCATCGGCGACTACAATATCGGTGGTGACGCCTGGTGTCCCGTATCCT
CCTGGAGGAGATGGGCCTGCGCGTGATCGCCAGTGGTCCGGCGACGGTTCGCTCGCTGAG
CTGGAGGCCACCCGAAGGCCAAGCTCAACGTGCTGCACTGCTACCGCTCCATGAACATACAT
CTCGCGCCACATGGAAGAGAAGTACGGTATCCCGTGGTGCGAGTACAACCTTCTCGGTCTTC
CAAGATCGCCGAGTCCCTGCGCAAGATCGCCAGCTACTTCGACGACAAGATCAAGGAAGGCG
30 CGGAGCGCGTCATCGCCAAGTATCAGCCGCTCATGGATGCGGTGATCGCGAAGTATCGTCCC
CGCCTCGAGGGCAAGACCGTGATGCTGTACGTGGGCGGCCTGCGTCCCCGTACGTCATCG
GCGCCTACGAGGACCTGGGCATGGAAGTGGTCGGCACGGGCTACGAGTTCGCCCATAACGA
CGACTACCAGCGCACCGCCAGCACTACGTCAAGGATGGCACCATCATCTATGACGACGTGA
CCGGCTACGAGTTCGAGAAGTTCGTGAGAAGATCCAGCCGGACCTGGTTCGGTTCGGGCATC
35 AAGGAAAAGTACGTCTTCCAGAAGATGGGCGTGCCGTTCCGCCAGATGCACTCCTGGGACTA
CTCGGGCCCCGTACCACGGCTATGACGGCTTCGCGATCTTCGCGCGCGACATGGACATGGCCA

TCAACAGCCCCGTGTGGAAGATGACCCAGGCTCCGTGGAAGAGCGTCCCCAAGCCGACGATG
CTCGCGGCTGAATGA

SEQ ID NO: 53:

- 5 Amino acid sequence of Nitrogenase molybdenum-iron protein alpha chain nifD_2
MSVAQSQSVAEIKARNKELIEEV LKVYPEKTAKRRAKHLNVHEAGKSDCGVKSNIKSIPGVMTIR
GCAYAGSKGVVWGPIKDMIHISHGPVGCQYSWAARRNYYIGTTGIDTFVTMQFTSDFQEKDIV
FGGDKKLAKIMDEIQELFPLNNGITVQSEPIGLIGDDIEAVSKQKSKEYEGKTIVPVRCEGFRGV
S QSLGHHIANDAI RDWVFDKIAPDAEPRFEPTPYDVAIIGDYNIGGDAWSSRILLEEMGLRVIAQ
10 WSGDGS LAELEATPKAKLNLH CYRSMNYISRHMEEKY GIPWCEYNFFGPSKIAESLRKIASYFD
DKIKEGAERVI AKYQPLMDAVIAKYRPRLEGKTVMLYVGGLRPRHVIGAYEDLGMEVVGTGYEFA
HND DYQRTAQHYVKDGTIIYDDVTGYEF EKFEKIQPDLVGSIGKEKYVFQKMGVPFRQMHSWD
YSGPYHGYDGF AIFARDMDMAINSPVWKMTQAPWKSVPKPTMLAAE

15 SEQ ID NO: 54:

- Nucleotide sequence of Nitrogenase molybdenum-iron protein beta chain nifK_1
ATGGCCACCGTTTCCGTCTCCAAGAAGGCCTGCGCGGTCAACCCCTCAAGATGAGCCAGCC
GGTGGGCGGCGCGCTCGCCTTCATGGGCGTGCGCAAGGCCATGCCGCTGCTGCACGGCTCG
CAGGGCTGCACCTCCTTCGGCCTGGTGCTGTTCTGCGCCACTTCAAGGAAGCCATCCCCAT
20 GCAGACCACCGCCATGAGCGAGGTGGCGACGGTTCTGGGCGGCCTTGAGAATGTGGAGCAG
GCCATTCTCAACATCTACAATCGCACCAAGCCGGAGATCATCGGCATCTGCTCCACCGGCGTC
ACCGAGACCAAGGGCGATGATGTGACGGCTACATCAAGCTGATCCGGGACAAGTATCCCCA
GCTGGCCGACTTCCCGCTGGTCTATGTCTCCACCCCGATTCAAGGACGCCTTCCAGGACG
GTTGGGAGAAGACCGTGGCGAAGATGGTGGAGGCGCTGGTGAAGCCCGCCGCGACAAGCA
25 GAAGGACAAGACCCGCGTCAACGTCTGCCCGGTGCCACCTCACGCCCGCGATCTGGATG
AGATGCGGACCATCTTCGAGGATTTCTGGGCTCACACCCTATTTCTGCCGGATCTGGCCGGCT
CGCTGGATGGGCATATCCCCGAGGACTTCTCGCCACCACCATCGGCGGCATCGGCATCGAT
GAGATCGCCACCATGGGCGAGGCGGCCACACCATCTGCATCGGCGCGCAGATGCGCCGGG
CGGGCGAGGCCATGGAGAAGAAGACCGGCATTCCCTTCAAGCTGTTTCGAGCGCCTGTGCGG
30 CCTGGAGGCGAACGACGCCTTCATCATGCACCTGTGCGAGATCTCCGGCCGGCCGGTGCCGG
TGAAGTATCGCCGGCAGCGGGGCCAGCTGGTGGATGCCATGCTGGACGGCCACTTCCATCTG
GGCGGTGCGAAGGTGGCCATGGGGGCGGAGCCGGACCTGCTCTACGACGTGGGCTCCTTCC
TGCACGAGATGGGCGCCACATCCTTCCGCGGTCAACCACACCCAGTCGCCGGTGCTGGCG
CGCCTGCCTGCCGAGGAGGTGCTTATCGGCGACCTGGAGGATCTGGAGACCCAGGCGAAGG
35 CGCGCGGATGCGATCTCCTGCTCACCCATTCCCATGGGCGCCAGGCGGCGGAGCGCCTCCAC
ATCCCCCTTACCGGATCGGCATTCCCATGTTTGACCGGCTGGGGGCGGGGCATCTGTTGTC
GGTGGGCTATCGCGGCACCCGCGACCTCATCTTCCATCTCGCCAACCTTGTGATCGCCGACCA

CGAGGAAAATCACGAGCCGACGCCCCGACACCTGGGCCACCGGCCATGGCGAGCATGCCGCC
GCCCCCACTTCCCATTGA

SEQ ID NO: 55:

- 5 Amino acid sequence of Nitrogenase molybdenum-iron protein beta chain nifK_1
MATVSVSKKACAVNPLKMSQPVGGALAFMGVRKAMPLHGSQGCTSFGLVLFVRHFKEAIPMQT
TAMSEVATVLGGLENVEQAILNIYNRTKPEIIGICSTGVTETKGDDVDGYIKLIRDKYPQLADFLV
YVSTPDFKDAFQDGWEKTVAKMVEALVKPAADKQKDKTRVNVLPGCHLTPGDLDDEMRTIFEDFG
LTPYFLPDLAGSLDGHIPEDFSPTTIGGIGIDEIATMGEEAHTICIGAQMRRAGEAMEKKTGIPFKL
10 FERLCGLEANDAFIMHLSQISGRPVVPKYRRQRGQLVDAMLDGHFHLGGRKVMGAEPDLLYDV
GSFLHEMGAHILSAVTTTQSPVLARLPAAEVLIGDLEDLETQAKARGCDLLLTHSHGRQAAERLHI
PFYRIGIPMFDRLGAGHLLSVGYRGTRDLIFHLANLVIADHEENHEPTPDTWATGHGEHAAAPT
H

- 15 SEQ ID NO: 56:

Nucleotide sequence of Nitrogenase molybdenum-iron protein beta chain nifK_2
ATGCCACAAAATGCTGACAATGTGCTCGATCACTTCGAGCTCTCCGTGGTCCCGAATACCAG
CAGATGCTGGCCAATAAGAAAAAGATGTTGAGAACCCCCGCGATCCGGCCGAAGTCGAGCG
CGTGCGGGAATGGGCGAAGACTCCTGAATACAAGGAGCTGAACTTCGCCCCGCGAGGCGCTC
20 ACCGTGAATCCGGCCAAGGCTTGTGACCCGCTGGGCGCGGTGTTCTGTCGCCGTGCGCTTCGA
GAGCACGATCCCCCTTCGTGCACGGCTCGCAGGGTTGCGTCGCGTATTACCGCTCGCACCTCT
CCCCCACTTCAAGGAGCCGTCTCTGCGTCTCCTCGTCCATGACCGAGGATGCGGCGGTG
TTCGGCGGCCTCAACAACATGATTGACGGCCTCGCCAACACCTACAACATGTACAAGCCGAAG
ATGATCGCCGTCTCCACCACCTGCATGGCGGAAGTCATCGGCGACGATCTGAACGCCTTCATC
25 AAGACCGCGAAGGAAAAGGGCTCGGTTCCGGCCGAATACGACGTGCCCTTCGCCCACACCCC
GGCGTTCTGTCGGCAGCCATGTCACCGGCTACGACAATGCGCTCAAGGGCATCCTCGAGCACT
TCTGGGACGGCAAGGCCGGCACCGCGCCGAAGCTGGAGCGCGTTCCTCAACGAGAAGATCAA
CTTCATCGGCGGCTTCGACGGCTACACCGTCGGCAACACTCGCGAAGTGAAGCGCATCTTCG
AGGCGTTCGGCGCCGATTACACCATCCTCGCCGACAATTCCGAAGTGTTGACACCCCCGACC
30 GACGGCGAGTTCGCGATGTATGACGGCGGCACGACCCTGGAGGACGCGGCGAAGCGGGTGC
ACGCCAAGGCCACCATCTCCATGCAGGAATACTGCACGGAGAAGACCTGCCCATGATCGCC
GGTCATGGCCAGGACGTGGTCGCCCTCAACCACCCCGTGGGCGTGGGCGGCACCGACAAGT
TCCTCATGGAGATCGCCCGCTCACC GGCAAGGAGATCCCCGAGGAGCTGACCCGCGAGCG
CGGCCGTCTCGTGGACGCTATCGCGGACTCTTCGCGCACATCCACGGCAAGAAGTTCGCCA
35 TCTACGGCGATCCGGATCTGTGCCTGGGCCTCGCCGCGTTCCTGCTGGAGCTGGGCGCCGAG
CCGACCCATGTGCTGGCCACCAACGGCACCAAGAAGTGGGCCGAGAAGGTTAGGAACTGTT
CGACTCTTCGCCGTTCGGCGCCAAGTCAAGGTCTATCCCGGCAAGGACCTGTGGCACATGC

GCTCGCTCCTGTTCTGTGGAGCCGGTGGATTTTCATCATCGGCAACACCTACGGCAAGTATCTCG
AGCGCGACACGGGCACCCCGCTGATCCGTATCGGCTTCCCGGTGTTTCGACCGTCACCACCAC
CACCGCCGTCCGGTGTGGGGCTATCAGGGCGGCATGAACGTCCTGATCACGATCCTCGACAA
GATCTTTGACGAGATCGACCGCAACACCAACGTGCCGGCCAAGACCGACTACTCGTTCGACAT
5 CATTCGTTGA

SEQ ID NO: 57:

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

MPQNADNVLDHFELFRGPEYQQMLANKKKMFENPRDPAEVERVREWAKTPEYKELNFAREALTV
10 NPAKACQPLGAVFVAVGFESTIPFVHGSQGCVAYYRSHLSRHFKEPSSCVSSSMTEAAVFGGLN
NMIDGLANTYNMYKPKMIAVSTTCMAEVIGDDLNAFIKTAKEKGSVPAEYDVPFAHTPAFVGSHV
TGYDNALKGILEHFWDGKAGTAPKLERPNEKINFIGGFDGYTVGNTREVKRIFEAFGADYILAD
NSEVFDTPTDGEFRMYDGGTTLEDAANAVHAKATISMQEYCTEKLPMIAGHGQDVVALNHPVG
VGGTDKFLMEIARLTGKEIPELTRERGRVLDAIADSSAHIHGKKFAIYGDPLCLGLAAFLLELGA
15 EPTHVLATNGTKKWAQKVQELFDSSPFGANCKVYPGKDLWHMRSLLFVEPVDFIIGNTYGKYLER
DTGTPLIRIGFPVDRHHHHRRPVWGYQGGMNVLITLDKIFDEIDRNTNVPKTDYSFDIIR

SEQ ID NO: 58:

Nucleotide sequence of Nitrogenase iron protein nifH

20 GTGGAGTCCGGTGGTCCTGAGCCGGGCGTGGGCTGCGCCGGCCGCGCGTGTACCTCCA
TCAACTTCCTGGAGGAGAACGGCGCCTACGAGGACATCGACTATGTGTCTACGACGTGCTG
GGCGACGTGGTGTGCGGCGGCTTCGCCATGCCATCCGCGAGAACAAGGCGCAGGAAATCT
ACATCGTGTATGTCGGCGAGATGATGGCCATGTATGCGGCCAACAAACATCTCCAAGGGCATC
CTGAAGTATGCCAATTCCGGCGGCGTGCCTGCGGCGGGCTGGTCTGCAACGAGCGCCAGA
25 CCGACAAGGAGCTGGAGCTGGCGGAGGCTCTGGCGAAGAAGCTCGGCACCGAGCTGATCTA
CTTCGTGCCGCGCGACAACATCGTGCAGCATGCCGAGCTGCGCCGCATGACAGTATCGAGT
ATGCGCCCGATTCCGCCCAGGCCAGCACTACCGGAACCTGGCCGAGAAGGTGCACGCCAAC
AAGGGCAACGGCATCATCCCGACCCCGATCACCATGGACGAGCTGGAAGACATGCTCATGGA
GCACGGCATCATGAAGGCCGTGGACGAGAGCCAGATCGGCAAGACCGCCGCGGAGCTCGCC
30 GTCTGA

SEQ ID NO: 59:

Amino acid sequence of Nitrogenase iron protein nifH

MESGGPEPGVGCAGRGVITSINFLEENGAYEDIDYVSYDVLGDVVCGGFAMPIRENKAQEIIYVM
35 SGEMMAMYAANNISKILKYANSGGVRLGGLVCNERQTDKELELAELAKKLGTELIYFVPRDNI
VQHAELRRMTVIEYAPDSAQAQHYRNLAIEKVHANKNGIIPITMDELEDMLMEHGIMKAVDES
QIGKTAELAV

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. 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, 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 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, resulting in a protein fraction and a fraction comprising non-protein components. Thus, in one embodiment, the process is for the production of protein and comprises a step of culturing strain VTT-E-193585 or a derivative thereof, followed by a step of harvesting biomass and a further step of isolating proteins from said biomass. In another embodiment, the process is for the production of protein and comprises culturing a bacterial strain of the genus *Xanthobacter* in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide, followed by a step of harvesting biomass and a further step of isolating proteins from said biomass. 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 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

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
5 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
10 "Technology of production of edible flours and protein products from soybeans" by Berk 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 further purified from this fraction using any suitable method.

15 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.

In one embodiment of the process of the invention, the process comprises the further
20 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 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
25 or fraction thereof is performed during the course of its incorporation into a food or feed product.

In a further aspect, the invention relates to a product, such as biomass, protein, or non-protein components, obtained or obtainable by the process according to the invention.

30

In one embodiment, the product obtained from the process of the invention comprises more than 40% protein, such as between 40% and 99% protein, e.g. between 40% and 90% protein, such as between 40% and 60% protein. In a particular embodiment, the product comprises between 25% and 75% protein, between 0% and 20% lipid and
35 between 5% and 40% carbohydrates. In a further embodiment, the product comprises between 40% and 60% protein, between 0% and 15% lipid and between 10% and 25% carbohydrate. In an even further embodiment, the product obtained from the process

of the invention comprises between 45% and 55% protein, between 5% and 10% lipid and between 10% and 20% carbohydrates.

As described above, the invention in a further aspect relates to a food or feed product obtained or obtainable by the process according to the invention. When used herein, the terms "food" and "feed" are intended to include not only conventional food and feed products, such as processed foods, but also related products, such as food and feed supplements, e.g. protein bars, powders or shakes, meat replacements, food ingredients, probiotics, prebiotics, nutraceuticals and the like. In certain embodiments, said biomass, said protein fraction or said fraction comprising non-protein components is utilized in the production of a vegetarian or vegan food product.

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

15 **EXAMPLES**

Example 1. Isolation of bacterial strain capable of chemoautotrophic growth

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 $\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$, 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 N_2 , 18 mL/min of H_2 , 3 mL/min of O_2 and 6 mL/min of CO_2 . 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. 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 (OD_{600}) 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 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 H_2 , 3 mL/min O_2 and 6 mL/min CO_2 . 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 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 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 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.

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 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.

A search for putative antimicrobial resistance genes was performed. The ABRicate (https://github.com/tseemann/abricate) 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.

Example 2. Pilot cultivation and analysis of isolated bacterial strain

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 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

5 **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

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

5 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 isolated 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, $\mu\text{g/ml}$) -values of antibiotics for VTT-E-193585 strain and *Escherichia coli* ATCC 25922

	VTT-E-193585 48 h $\pm$ 1 h	<i>E. coli</i> ATCC 25922 18 h $\pm$ 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. An isolated bacterial strain VTT-E-193585 or a derivative thereof, wherein said derivative preferably has retained the ability to grow using hydrogen gas as energy source and carbon dioxide as the only carbon source.
5
2. A culture comprising the bacterial strain according to claim 1.
3. A process for the production of biomass, said process comprising culturing the bacterial strain of claim 1.
10
4. The process according to claim 3, comprising culturing the bacterial strain in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide.
15
5. A process for the production of biomass, said process comprising culturing a bacterial strain of the genus *Xanthobacter* in continuous culture with hydrogen as energy source and an inorganic carbon source, wherein the inorganic carbon source comprises carbon dioxide.
20
6. The process according to claim 4 or 5, wherein dissolved oxygen in the culture is maintained between 5% and 10%.
7. The process according to any one of claims 4 to 6, wherein ammonium, urea, nitrate and/or nitrogen gas is used as nitrogen source.
25
8. The process according to any of claims 4 to 7, wherein the culture medium comprises minerals wherein said minerals 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, preferably no chloride salts.
30
9. The process according to any one of claims 4 to 8, wherein the culture medium does not contain vitamins.
- 35 10. The processing according to any one of claims 4 to 9, wherein 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.

11. The process according to any one of claims 4 to 10, wherein said 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.

5 12. The process according to any one of claims 4 to 11, wherein the bacterial strain has a growth rate of 0.04-0.12 h⁻¹.

13. The process according to any one of claims 4 to 12, wherein the bacterial strain is VTT-E-193585 or *X. taetidis*.

10

14. The process according to any one of claims 4 to 13, 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.

15

15. The process according to any one of claims 4 to 14, 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 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, 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, 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, and/or

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, 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: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 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, and/or

5 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, 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:11, and/or

10 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, 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
15 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, 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
20 identity to 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, 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
25 NO:17, and/or

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, 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
30 more than 99% sequence identity to the sequence set forth in SEQ ID NO:19, and/or

x) 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
35 99% sequence 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, 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:23, and/or

- 5 xii) a gene encoding an ATP synthase subunit c, sodium ion specific atpE_1 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, and/or
- 10 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, 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:27, and/or
- 15 xiv) 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 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, and/or
- 20 xv) 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, and/or
- 25 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, e.g. more than 98%, such as more than 99% sequence identity to 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, 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:35, and/or
- 30 xviii) 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 99% sequence identity to the sequence set forth in SEQ ID NO:37, and/or
- 35

- 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, 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:39, and/or
- 5 xx) 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, and/or
- 10 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, 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, and/or
- 15 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%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:45, and/or
- 20 xxiii) 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, and/or
- 25 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, 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, and/or
- 30 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, 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. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID
- 35 NO:51, and/or
- 36 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,

- e.g. more than 98%, such as more than 99% sequence identity to the sequence set forth in SEQ ID NO:53, and/or
- 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, 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, 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 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: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),
 - the genes described in vii) and viii), or
 - the genes described in iii), iv), v), vi), vii) and viii).
16. The process according to any one of claims 4 to 15, comprising the further step of harvesting biomass produced during the culture, optionally comprising a further step of drying the biomass.
17. A process for the production of protein, comprising performing the process according to claim 16 and a further step of isolating protein from said biomass, wherein the process results in a protein fraction and a fraction comprising non-protein components.
18. The process according to claim 16 or 17, 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.
19. A product, such as biomass, protein, or non-protein components obtained or obtainable by the process according to any one of claims 4 to 18.
20. A food or feed product obtained or obtainable by the process according to claim 18.