Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria

Paula B. Matheus Carnevali1, Frederik Schulz2, Cindy J. Castelle1, Rose S. Kantor1,13, Patrick M. Shih2,3,4,14, Itai Sharon1,15,16, Joanne M. Santini5, Matthew R. Olm6, Yuki Amano7,8, Brian C. Thomas1, Karthik Anantharaman1,17, David Burstein1,18, Eric D. Becraft19,9, Ramunas Stepanauskas9, Tanja Woyke2 & Jillian F. Banfield1,6,10,11,12

The evolution of aerobic respiration was likely linked to the origins of oxygenic Cyanobacteria. Close phylogenetic neighbors to Cyanobacteria, such as Margulisbacteria (RBX-1 and ZB3), Saganbacteria (WOR-1), Melainabacteria and Sericytochromatia, may constrain the metabolic platform in which aerobic respiration arose. Here, we analyze genomic sequences and predict that sediment-associated Margulisbacteria have a fermentation-based metabolism featuring a variety of hydrogenases, a streamlined nitrogenase, and electron bifurcating complexes involved in cycling of reducing equivalents. The genomes of ocean-associated Margulisbacteria encode an electron transport chain that may support aerobic growth. Some Saganbacteria genomes encode various hydrogenases, and others may be able to use O2 under certain conditions via a putative novel type of heme copper O2 reductase. Similarly, Melainabacteria have diverse energy metabolisms and are capable of fermentation and aerobic or anaerobic respiration. The ancestor of all these groups may have been an anaerobe in which fermentation and H2 metabolism were central metabolic features. The ability to use O2 as a terminal electron acceptor must have been subsequently acquired by these lineages.
Oxygenic photosynthesis evolved in Cyanobacteria and ultimately led to the Great Oxidation Event (GOE) ~2.3 billion years ago. This was a major step in the co-evolution of life and the planet. The acquisition of the combination of photosystem I and photosystem II by Cyanobacteria allowed exploitation of water (H₂O) as an electron donor and the production of oxygen (O₂). The transport of electrons through the chain of photosystem complexes enables export of protons (H⁺) and generation of a proton motive force (PMF) that can be harnessed to form adenosine triphosphate—ATP). Formation of ATP is described as energy conservation because ATP is required for reactions such as carbon dioxide (CO₂) fixation, nitrogen (N₂) fixation and biosynthesis. Constitutive expression of fermentation pathways in Cyanobacteria suggest that prior to the advent of oxygenic photosynthesis, the ancestral mechanism for ATP formation in these organisms was fermentation, which does not require a separate electron acceptor.

Following the GOE, organisms had access to higher energetic yield from aerobic respiratory processes that involve coupling oxidation of an electron donor such as organic carbon to reduction of O₂. Also following the GOE, the potential for anaerobic respiration would have greatly increased due to the availability of oxidized compounds, such as nitrate (NO₃⁻) that can also serve as an electron acceptor (NO₃⁻ forms in the environment mostly via O₂-dependent reactions). Aerobic respiration and NO₃⁻ reduction required the evolution of redox complexes of the electron transport chain (ETC), and use of O₂ required the evolution of a terminal oxidase. The simplest ETC is composed of (i) an electron entry point, usually a dehydrogenase that oxidizes reduced electron carriers (e.g., reduced nicotinamide adenine nucleotide – NADH and succinate), (ii) membrane electron carriers (e.g., quinones), and (iii) an electron exit point such as a heme-copper oxygen reductase or a cytochrome bd oxidase. However, if there are periplasmic electron donors, an additional electron entry point such as a heme-copper oxygen reductase or a cytochrome bd oxidase may also be involved. Both the electron entry and exit protein complexes often act as H⁺ pumps, contributing to the creation of a PMF.

The metabolic potential of lineages phylogenetically related to Cyanobacteria is of great interest from the perspective of constraining the biological context in which complex ETCs evolved. Recently, several publications have investigated the biology of Melainabacteria and Sericytochromatia, groups that branch adjacent to Cyanobacteria and are represented by non-photosynthetic organisms. It has been suggested that the machinery for aerobic respiration was acquired independently by Sericytochromatia, and like Cyanobacteria, some members of the Sericytochromatia and Melainabacteria have ETCs. Analysis of genomes of additional major groups of bacteria sibling to the Cyanobacteria may help distinguish the possibilities that Melainabacteria and Sericytochromatia acquired anaerobic metabolisms after their divergence from Cyanobacteria from the alternative, in which Cyanobacteria gained aerobic metabolism after their divergence from Melainabacteria and Sericytochromatia.

Here we investigate genomes of Margulisbacteria (RBX-1) and Saganbacteria (WOR-1), lineages related to both Melainabacteria and Sericytochromatia, and identify common mechanisms for energy conservation that may have been present in their common ancestor with Cyanobacteria. Several genomes were previously reconstructed from estuarine sediments and groundwater. For Margulisbacteria, we describe two groups that we refer to as Riflemargulisbacteria (given the derivation of draft genome sequences from the Rifle research site), and four groups we refer to as Marinamargulisbacteria (given the derivation of the single-cell draft genomes from ocean sites). Marinamargulisbacteria was first observed in a clone library of SSU rRNA genes from Zodletone spring (Oklahoma) and identified as candidate division ZB3. We leverage these genomes, and new genomes for Margulisbacteria and members of the Melainabacteria to predict how organisms from these lineages conserve energy in the form of ATP, identify their potential strategies for reoxidation of reducing equivalents, and propose the interconnectedness between H₂ and O₂ metabolism. Membrane-bound complexes that may be involved in ion translocation for generation of an electrochemical membrane potential were of particular interest. We found that H₂ metabolism is a common feature of the modern groups that cannot respire aerobically, and suggest that hydrogenases may have been central to the lifestyles of the ancestors of Cyanobacteria, Melainabacteria, Sericytochromatia, Margulisbacteria, and Saganbacteria.

**Results**

**Identification of representative genomes from each lineage.** In this study, we analyzed four publicly available and eight newly reconstructed genomes of Margulisbacteria and propose two distinct clades within this lineage: Riflemargulisbacteria (two groups) and Marinamargulisbacteria (four groups). These clades were defined using metagenome-assembled genomes (MAGs) of bacteria from the sediments of an aquifer adjacent to the Colorado River, Rifle, USA and single-cell amplified genomes (SAGs) from four ocean environments (Supplementary Data 1). We also analyzed 26 publicly available MAGs of Saganbacteria from groundwater with variable O₂ concentrations, of which four genomes are circularized. In addition, we analyzed eight publicly available MAGs of Melainabacteria, and five new Melainabacteria genomes that were reconstructed from metagenomic datasets for microbial communities sampled from human gut and groundwater (Supplementary Data 1). We identified representative genomes for groups of genomes that share 95.0–99.0% average nucleotide identity (ANI) (Supplementary Data 1) and focus our discussion on these genomes. Metabolic predictions for genomes that were estimated to be medium or high quality were established using gene annotations and confirmed using Hidden Markov Models (HMMs) built from the KEGG database (Supplementary Data 2). Key aspects of energy metabolism predicted for Riflemargulisbacteria, Marinamargulisbacteria, Saganbacteria, and Melainabacteria were compared. Phylogenetic analyses of genes encoding key metabolic functions also included publicly available genomes for Sericytochromatia, newly reported genomes for relatives of Marinamargulisbacteria, and other reference genomes.

**Relatives of Cyanobacteria, Melainabacteria, and Sericytochromatia.** In our phylogenetic analysis, Margulisbacteria and Saganbacteria group together in sibling position to a monophyletic clade consisting of Sericytochromatia, Melainabacteria, and Cyanobacteria (Fig. 1a, b). In addition, Margulisbacteria, Saganbacteria and Sericytochromatia are positioned basally to Melainabacteria and Cyanobacteria. Importantly, the affiliation of these groups with the Cyanobacteria is consistent for phylogenetic trees based on 56 universal single copy proteins and 16 ribosomal proteins, which are a subset of the 56 universal single copy proteins (Fig. 1a, Supplementary Fig. 1, 2). However, the precise position of these three lineages in the Terrabacteria differs depending on the set of markers and phylogenetic models used for tree construction (Fig. 1a, Supplementary Fig. 2) and the deep branching order often cannot be well resolved. In addition, the position of Margulisbacteria, Saganbacteria, and Cyanobacteria relative to the Thermoanaerobacterales (Fig. 1a)
H₂ metabolism may be vital for generating a proton-motive force. Metabolic analyses based on a representative genome of the sediment-associated Riflemargulisbacteria, Margulisbacteria RA1A (96% completeness, representative of four genomes in the same ANI cluster, see Supplementary Data 1), suggest that members of Riflemargulisbacteria are heterotrophic organisms with a fermentation-based metabolism (Fig. 2). They use hydrogenases and other electron bifurcating complexes to balance reducing equivalent pools (NADH and ferredoxin), and rely on membrane-bound protein complexes to generate a proton/sodium (H⁺/Na⁺) potential and to make ATP (Fig. 3a). We predict that Margulisbacteria RA1A is an obligate anaerobe, and multiple enzymes that participate in central metabolism use ferredoxin as the preferred electron donor/acceptor (Fig. 2). Furthermore, Margulisbacteria RA1A lacks most components of the tricarboxylic acid (TCA) cycle and an ETC (Supplementary Data 2), including terminal reductases for aerobic or anaerobic respiration. Carbon dioxide (CO₂) fixation pathways were not identified in genomes of this lineage (Supplementary Data 2). For further details about the central metabolism, lifestyle, and other features of Riflemargulisbacteria see Supplementary Notes 1, 2, 7 and 8 and Supplementary Data 3 and 4.

Hydrogen metabolism appears essential for the sediment-associated Riflemargulisbacteria. In the absence of a well-defined respiratory electron transport chain, reduced ferredoxin and NADH can be re-oxidized by reducing H₂ to hydrogen (H₂). Hydrogenases, the key enzymes in hydrogen metabolism, can be reversible and either use H₂ as a source of reducing power or H₂ as oxidants to dispose of excess reducing equivalents. The assignments of hydrogenase types in this organism were based on phylogenetic analysis of the hydrogenase catalytic subunits (Fig. 4 and Supplementary Fig. 4), analysis of the subunit composition and predicted protein domains. Notably, only NiFe hydrogenases...
oxidoreductase, ADH glutamine synthetase, oxidoreductase (Nuo), a protein complex that couples NADH bound group 4 NiFe hydrogenase, and a hydrogenase-related type of cytoplasmic FeFe hydrogenase (Supplementary Fig. 4 and potentially membrane-bound complexes of Ri. potential in prokaryotic cells, we focused our attention on the bound protein complexes in the generation of a membrane Data 8). For a detailed description of cytoplasmic hydrogenases from groups 3 and 4 are encoded in the genome, and these seem to have been horizontally transferred based on phylogenetic analyses. Group 4 NiFe hydrogenases are thought to be among the most primitive respiratory complexes.

We identified three types of cytoplasmic NiFe hydrogenases (groups 3b, 3c, and 3d; Fig. 4a; Supplementary Data 5-8) and one type of cytoplasmic FeFe hydrogenase (Supplementary Fig. 4 and Supplementary Data 9-12), in addition to one type of membrane-bound group 4 NiFe hydrogenase, and a hydrogenase-related complex (Fig. 4b; Supplementary Data 13-15 and Supplementary Data 8). For a detailed description of cytoplasmic hydrogenases see Supplementary Note 3. Given the importance of membrane-bound protein complexes in the generation of a membrane potential in prokaryotic cells, we focused our attention on the potentially membrane-bound complexes of Rilemargulisbacteria. Membrane-bound group 4 NiFe hydrogenases and related proteins share a common ancestor with NADH:ubiquinone oxidoreductase (Nuo), a protein complex that couples NADH oxidation with H⁺ or ion translocation. This complex is a key component of respiratory and photosynthetic electron transport chains in other organisms.

The Rilemargulisbacteria are predicted to have a group 4 membrane-bound NiFe hydrogenase (Fig. 4b), also called energy-converting hydrogenase (Ech). These enzymes usually couple the oxidation of reduced ferredoxin to H₂ evolution under anaerobic conditions, can be reversible, and are typically found in methanogens and other anaerobic organisms where they function as primary proton pumps. Genes for two other complexes are found in close proximity to the Ech-type hydrogenase, a V-type ATPase and a Rhodobacter nitrogen fixation (Rnf) electron transport complex (Fig. 3a). These protein complexes may also contribute to the generation of an electrochemical potential in Margulisbacteria RA1A. This is remarkable because both the Ech hydrogenase and the Rnf complex are capable of generating a transmembrane potential, and only a small number of bacteria have been observed to have both. The putative V-type ATPase operon in Margulisbacteria RA1A (atpEXABDIK) resembles those of Spirochetes and Chlamydiales (Supplementary Fig. 5).
Data 16). This complex could hydrolyze or make ATP with $H^+$ or Na$^+$ translocation out of the cell. Likewise, the Rnf complex, which generally functions as a ferredoxin:NAD$^+$ oxidoreductase in *Rhodobacter capsulatum*, can couple an exergonic redox reaction to Na$^+$ translocation across the membrane24. The Rnf operon in Margulisbacteria RA1A is composed of six genes (*rnfCDGEAB*), similar in organization to the operon in *Acetobacterium woodii* and evolutionarily related to the Na$^+$-translocating NADH:quinone oxidoreductase (Nqr)24,25.

Ehr complexes are suspected to couple catalytic oxidation-reduction reactions in the cytoplasm to charge translocation across the membrane16, due to the presence of multiple resistance and pH adaptation (Mrp) antiporter-like subunits16. For instance, in *Pyrococcus furiosus* a membrane-bound oxidoreductase (Mb; a type of hydrogenase-related complex) functions as a ferredoxin: nicotinamide adenine dinucleotide phosphate (NADP$^+$) oxidoreductase complex that potentially pumps $H^+$ across the membrane27. Given the Ehr subunit composition in *R. fleremargulisbacteria* (Supplementary Fig. 3a and Supplementary Note 4) and the predicted need for an electron donor component, we looked for potential partners. In *R. fleremargulisbacteria*, the candidate cytoplasmic oxidoreductase partner may be a molybdenum- or tungstate-dependent formate dehydrogenase (*fdhA*; Fig. 5; Supplementary Data 17), which likely oxidizes formate to CO$_2$ and NADPH, possibly reversibly. That NAD(P)$^+$ may be the electron acceptor for this reaction is inferred from the presence of two genes (*nuoE* and *nuoF*-like) next to *fdhA*, encoding FeS cluster-binding domains and an electron carrier flavin mononucleotide (FMN)-binding domain (Supplementary Fig. 3b and Supplementary Note 4). Therefore, we hypothesize the presence of a formate dehydrogenase (FDH)-Ehr complex. Notably, genes encoding a full pathway for Na$^+$/H$^+$ antiporter-like subunits.
Fig. 4 Phylogenetic tree of the catalytic subunit in NiFe hydrogenases. a Bayesian phylogeny indicating the phylogenetic relationships of the Rijemargulisbacteria, Saganbacteria, and Melainabacteria groups 1, 2 and 3 NiFe hydrogenase catalytic subunits and b Bayesian phylogeny indicating the phylogenetic relationships of the Rijemargulisbacteria, Saganbacteria, and Melainabacteria group 4 NiFe hydrogenases and related complexes. Scale bar indicates substitutions per site. Branches with a posterior support of below 0.5 were collapsed. The tree and underlying alignments are available with full bootstrap values as pdf and in Newick format in Supplementary Data 5-8 and 13-15.
synthesis of ubiquinone and menaquinone, which could serve as the electron acceptor for the overall reaction catalyzed by such complex were not identified in these genomes.

Lastly, Margulisbacteria RA1A encodes a nitrogenase, an enzyme known to produce H₂ during nitrogen fixation. A phylogenetic analysis of key nitrogenase subunits NifHDK (Fig. 6; Supplementary Data 18-21) indicates that this enzyme is different from the nitrogenase in many Cyanobacteria. The nitrogenase operon (nifHI) in R. fleremargulisbacteria resembles that of methanogens. Besides genes encoding the Fe protein (NifH) and the MoFe protein (NifDK), this operon includes genes involved in post-translational regulation (nifI) and the synthesis of the enzyme (nifENBV). Additionally, nifA is found downstream from the operon, which is involved in transcriptional regulation. Therefore, this operon is of slightly higher complexity than the operon in methanogens (following). Phylogenetically, NifHDK in R. fleremargulisbacteria are related to Spirochetes, and other anaerobic taxa at the base of the tree (Fig. 6). Furthermore, we identified three genes that encode an FeFe hydrogenase (Supplementary Fig. 4) downstream of the nif operon. Given the proximity of the FeFe hydrogenase genes on the Margulisbacteria RA1A genome to the Nif operon, this enzyme may be responsible for dissipating excess H₂ generated by the nitrogenase complex involved in N₂ fixation and its expression could be regulated by nitrogen availability. Based on the predicted subunits domain composition, this enzyme could be classified as a group 2 (G2) bifurcating hydrogenase with a 3c modular structure. In the forward direction, FeFe hydrogenases can bifurcate electrons from H₂ to ferredoxin and nicotinamide adenine dinucleotide phosphate (NAD(P)H). Interestingly, FeFe hydrogenases have not been found in Cyanobacteria (or Saganbacteria).

Like Margulisbacteria, Saganbacteria have anaerobic and aerobic representatives. Similar to sediment-associated Margulisbacteria, Saganbacteria include heterotrophic anaerobes that may generate a H⁺/Na⁺ potential via membrane-bound group 4 NiFe hydrogenases and hydrogenase-related complexes (Fig. 4b), in combination with an Rnf complex.

The studied genomes of Saganbacteria encode distinct types of hydrogenases. In the category of cytoplasmic complexes, we only identified one type, the group 3b NiFe hydrogenases (sometimes referred to as sulfhydrogenases), and these occurred in multiple representative genomes (Fig. 4a, Supplementary Data 22). Many Saganbacteria representative genomes also encode three distinct types of membrane-bound hydrogenases (Fig. 4b). Genomic regions including fdhA and nuoE-and nuoF-like genes also occur in many Saganbacteria with Ehr complexes (Supplementary Figs. 3b and c), an observation that strengthens the deduction that an FDH-Ehr complex may occur in these clades.

Another type of group 4 NiFe hydrogenase found in Saganbacteria RX5A is enigmatic. The most closely related...
sequences are hydrogenases found in candidate phyla such as Zixibacteria and Omnitrophica (Fig. 4b, Supplementary Data 13). The genomic region encodes hydrogenase subunits, Mrp antiporter-like subunits, and a molybdopterin-containing oxidoreductase that could not be classified by phylogeny. Specifically, it could not be identified as a formate dehydrogenase, although the HMMs suggested this to be the case. Also present in the genomic region are a putative CO dehydrogenase subunit F gene (cooF), and putative anaerobic sulfite reductase subunits A (asrA) and B (asrB) genes (Supplementary Fig. 3e and Supplementary Note 5). Together, these subunits might compose an oxidoreductase module. We hypothesize that the function of the proteins encoded in this region is similar to that of Mbh hydrogenases described in other organisms.

Fig. 6 Phylogenetic tree of concatenated nitrogenase subunits NifHDK. Maximum likelihood tree indicating the branching of the Riflemargulisbacteria NifHDK together with anaerobic taxa. Scale bar indicates substitutions per site. Branches with a posterior support of below 0.5 were collapsed. The tree and underlying alignments are available with full bootstrap values as pdf and in Newick format in Supplementary Data 18-21.
Two other putative group 4 NiFe hydrogenases were found in Saganbacteria RX6A (Supplementary Figs. 3d and f). Each shares some components with the Mbh hydrogenase described above. One of them has Mrp antiporter-like subunits, but lacks an oxidoreductase module, and it is phylogenetically most closely related to hydrogenases in Pelobacter propionicus and Closstria (Fig. 4b, Supplementary Data 13 and Fig. 3f). Potentially this could be another Mrp-Mbh complex, and the lack of an oxidoreductase module combined with the presence of Fe–S cluster-binding domains indicates that it may be involved in ferredoxin reoxidation, as occurs in P. furiosus. The other putative group 4 NiFe hydrogenase harbors an oxidoreductase module, seems to lack Mrp antiporter-like subunits, and could not be placed phylogenetically due to a truncated sequence encoding a putative NiFe hydrogenase large (catalytic) subunit (Supplementary Fig. 3d). Therefore, we could not determine whether it is a Mbh-type hydrogenase or another type of hydrogenase-related complex (e.g. Mbx).

The majority of anaerobic Saganbacteria also possess genes encoding a Rnf complex and a V/A-type ATPase (Supplementary Data 2). The V/A-type ATPases have a subunit composition different from that in Riflemargulisbacteria, and their genes are in operons resembling those found in other bacteria and archaea (Supplementary Data 16).

Given the importance of hydrogenases and overall similarity in metabolic capacities of Riflemargulisbacteria and Saganbacteria, we investigated newly available genomes for the Melainabacteria for evidence of yet unrecognized H2 metabolism in this group. We found that some representative Melainabacteria genomes have bifurcating FeFe hydrogenases (modular groups 3a and 3c) that are monophyletic with the FeFe hydrogenase in Margulisbacteria, albeit in a separate phylogenetic cluster (Supplementary Fig. 4). For more detail about cytoplasmic hydrogenases in Melainabacteria, see Supplementary Note 6.

Ehr complexes are present in four representative Melainabacteria genomes (Fig. 4b). The sequences in the three of the four representative genomes cluster with sequences from Riflemargulisbacteria, Lentisphaerae, Spirochetes and Acinetobacter sp., but Melainabacteria BJ4A branches on its own (Fig. 4b). In addition to Ehr, Melainabacteria RX6A and BJ4A have an fdhA gene, and RX6A has genes encoding NuoE- and NuoF-like subunits just downstream of the fdhA gene. Interestingly, FdhA clusters with proteins found in the Cyanobacteria Nostoc piscinale and Scytomena hofmannii (Fig. 5). Melainabacteria HO7A is the only one with a putative group 4 f hydrogenase (after16), which is closely related to Gracilibacteria sp. and other anaerobic organisms.

Some Margulisbacteria RA1A, Saganbacteria and Melainabacteria have complexes predicted to be involved in electron bifurcation. For hydrogenase-based electron bifurcation, reduced ferredoxin and NAD(P)H (redox carriers with very different electrode potentials) are oxidized, coupled to the reduction of H+ to form H2.33 This reaction can also work in the reverse as the electron bifurcation complex can be bidirectional. Electron bifurcation complexes allow balance of reduced and oxidized cofactors in anaerobic environments, where electron acceptors are limited.

In close proximity to the region encoding FdhA in Margulisbacteria RA1A and most Saganbacteria genomes, we identified genes encoding a NADH-dependent reduced ferredoxin: NAD oxidoreductase (Nfn) complex (Fig. 3a). Nfn complexes use 2 NADPH molecules to catalyze the reversible reduction of oxidized ferredoxin and NAD+, thus it is involved in electron bifurcation.

Margulisbacteria RA1A, all Saganbacteria and many Melainabacteria have genes encoding a putative butyryl-CoA dehydrogenase (Bcd) in close proximity to genes encoding electron transfer flavoprotein subunits EtfA and EtfB (Fig. 3a). Together, these may form a Bcd/EtfAB complex, which is usually involved in electron bifurcation reactions between crotonyl-CoA, ferredoxin and NADH. Specifically, Bcd catalyzes the transformation of short-chain acyl-CoA compounds to short-chain trans-2,3-dehydroacetyl-CoA using electron-transfer flavoproteins as the electron acceptor (EtfAB). Nevertheless, the presence of an AMP-binding domain(s) instead of a second FAD-binding domain in the Etf indicates that electron bifurcation is not likely in this complex.36

Alternate mechanisms for the generation of a membrane potential. Margulisbacteria AA1A, the selected representative genome for the oceanic Marinamargulisbacteria (82% completeness, Supplementary Data 1) shows signs of niche adaptation, including the ability to use O2 as a terminal electron acceptor. This organism relies on aerobic respiration for generation of a H+ potential, and instead of pyruvate fermentation to short-chain fatty acids or alcohols, it may use acetate as a source of acetyl-CoA (Supplementary Data 2). Phylogenetic analysis of the heme-copper oxygen reductase (complex IV) indicated that the potential for high-energy metabolism was an independently acquired trait (via HGT) in this organism (Fig. 7, Supplementary Fig. 5).

Like the sediment-associated Margulisbacteria, the oceanic Marinamargulisbacteria lack CO2 fixation pathways. Thus, we anticipate that they adopt a heterotrophic lifestyle. The genome encodes all enzymes in the TCA cycle (including alpha-ketoglutarate dehydrogenase). NADH produced via glycolysis and the TCA cycle must be reoxidized. Unlike Riflemargulisbacteria that have membrane-bound NiFe hydrogenases, in Marinamargulisbacteria AA1A we identified genes encoding a six subunit Na+-translocating NADH:ubiquinone oxidoreductase (Nqr; EC 1.6.5.8; nqrABCDE; Supplementary Data 2). An electron transport chain in Marinamargulisbacteria AA1A (Fig. 3b) could involve electron transfer from NADH to the Nqr, from the Nqr to a menaquinone, then to a quinol:electron acceptor oxidoreductase (alternative Complex III; Fig. 5), and finally to a terminal oxidase (Fig. 7).

Phylogenetic analysis of the gene encoding the catalytic subunit (CoxA) of the terminal oxidase confirms that it is a type A heme-copper oxygen reductase (CoxABCD, EC 1.9.3.1; cytochrome c oxidase, (Fig. 7). The genes in these organisms are divergent and were probably acquired by horizontal gene transfer based on phylogenetic analysis (Supplementary Fig. 5). The recently released metagenome assembled Marinamargulisbacteria genome UBA6595 from Parks et al13 also encodes an alternative complex III and a type A heme-copper oxygen reductase. In contrast, the Riflemargulisbacteria do not have components of an ETC, except for a single gene of the cytochrome oxidase (coxB) presently with unknown function by itself (Supplementary Data 22). Interestingly, we did not identify any hydrogenases in the oceanic Marinamargulisbacteria, probably due to their different habitat, similar to many Cyanobacteria in the ocean.37

As occurs in the oceanic Margulisbacteria, aerobic organisms within the Saganbacteria also rely on an ETC to generate a H+ potential. Notably, based on phylogenetic analysis these aerobic organisms encode what looks like a novel type of heme-copper oxidase reductase (Fig. 7), and one organism (HO1A) also encodes the potential for anaerobic respiration (tetraionate reductase; Fig. 5).

Saganbacteria HO1A and LO2A possess genes encoding a partial Nuo complex I that lacks the NADH binding subunits (NuoEFG) and could use ferredoxin as the electron donor instead...
They also encode a succinate dehydrogenase (SDH) complex that is composed of four subunits and presumably fully functional as a complex II (Supplementary Data 2). Genes encoding a putative quinol:electron acceptor oxidoreductase complex, which has been previously seen in candidate phyla Zixibacteria 38, may transfer electrons to a terminal reductase and act as a complex III. Quinones, the usual electron donors for this and other respiratory enzymes, were not identified in this genome and alternative biosynthetic pathways should be investigated (Supplementary Data 2). Remarkably, these genomes also encode subunits of a quinol oxidase heme-copper oxygen reductase that seems to be a novel type, and that is closely related to a heme-copper oxygen reductase in Candidatus Methylomirabilis oxyfera (NC10 phylum) (Fig. 7) of unknown type. It is unlikely that these organisms can reoxidize NADH via complex I, and it remains uncertain whether they can use O2 as an electron acceptor. Downstream from the genes encoding the putative quinol reductase in Saganbacteria HO1A, there are genes encoding a tetrathionate reductase (Fig. 5), indicating that tetrathionate could be used as a terminal electron acceptor during anaerobic respiration, generating thiosulfate 39.

Melainabacteria include organisms capable of fermentation, respiration, or both5,6. In this study, Melainabacteria (except for AS1A and AS1B) have a partial complex I, similar to the one in Saganbacteria (Supplementary Data 22). They also encode a succinate dehydrogenase (SDH) complex that is composed of four subunits and presumably fully functional as a complex II (Supplementary Data 2). Genes encoding a putative quinol:electron acceptor oxidoreductase complex, which has been previously seen in candidate phyla Zixibacteria 38, may transfer electrons to a terminal reductase and act as a complex III. Quinones, the usual electron donors for this and other respiratory enzymes, were not identified in this genome and alternative biosynthetic pathways should be investigated (Supplementary Data 2).

Fig. 7 Phylogenetic tree of the catalytic subunit I of heme-copper oxygen reductases. The heme-copper oxygen reductase found in Saganbacteria seems to be a novel type based on phylogenetic analysis of catalytic subunits (CoxA). Marinamargulisbacteria have an A2 type heme-copper oxygen reductase, which is usually found in Cyanobacteria. The enzyme in Marinamargulisbacteria is part of a separate phylogenetic cluster from the A2-type heme-copper oxygen reductase found in Cyanobacteria (Supplementary Fig. 5). The heme-copper oxygen reductase found in Sericyctochromatia LSPB_72 is type A1, and CBMW_12 has two enzymes: one of type A1 and the other of type A2. Sericyctochromatia LSPB_72 and CBMW_12 also have type C heme-copper oxygen reductases (CcoN)7, like Melainabacteria in this study. Two types of bacterial nitric oxide reductase (NOR) have been identified in Sericyctochromatia. One is a cytochrome bc1-type complex (cNOR) that receives electrons from soluble redox protein donors, whereas the other type (qNOR) lacks the cytochrome c component and uses quinol as the electron donor. The tree is available with full bootstrap values in Newick format in Supplementary Data 25.
Discussion

Although billions of years have likely passed since the divergence of Cyanobacteria, Melainabacteria, Sericytchromatia, Margulisbacteria, and Saganbacteria, it is possible that features shared by the modern organisms were established prior to their divergence. Contemporary non-photosynthetic relatives of the Cyanobacteria share several interesting genomic features that frame their energy metabolism. These include numerous hydrogenases, lack of an oxidative phosphorylation pathway in many organisms, widespread lack of carbon dioxide fixation pathways, and the presence of multiple types of membrane-bound complexes that may be involved in ion translocation. Against this background, oxygenic photosynthesis and aerobic respiration represent a dramatic net-energy gain and access to a different life strategy—one that does not rely on limited organic carbon as a source of electrons, or complexes that provide only low-energy yields linked to ion translocation. Based on our analyses, we suggest that the common ancestor of all of these organisms was more likely an anaerobe and all but Cyanobacteria subsequently lost this capacity. Even today, Cyanobacteria can switch to fermentative metabolism under anoxic conditions, and at the heart of this process is a bidirectional group 3d NiFe hydrogenase \(^{41}\). Furthermore, the presence of a low complexity nitrogenase operon in Rilefamargulisbacteria suggests that this enzyme may have been present in anaerobic bacteria before the split between Cyanobacteria and Rilefamargulisbacteria.

Anaerobic representatives of Rilefamargulisbacteria and Saganbacteria have multiple protein complexes that may participate in the coupling of cytoplasmic redox reactions to ion translocation across the membrane. In Rilefamargulisbacteria and Saganbacteria, membrane-bound protein complexes (e.g., group 4 NiFe hydrogenases and hydrogenase-related complexes) could make up a minimal, highly efficient respiratory chain, as previously suggested for these types of enzymes. \(^{41}\) Sequential translocation of ions across membranes is required in fermentation-based metabolisms as the energy yield per translocation is low. \(^{41}\) Rilefamargulisbacteria is unique among the anaerobic lineages because its genome is predicted to encode multiple cytoplasmic H\(_2\)-consuming hydrogenases, in addition to membrane-bound energy-converting H\(_2\)-evolving NiFe hydrogenase (Ech), a Rnf complex, and potentially an FDH-Ehr complex. The cytoplasmic hydrogenases constitute the main mechanism by which these bacteria oxidize reducing equivalents that result from core metabolic pathways and produce H\(_2\). The main function of the Ech hydrogenase and the Rnf complex may be to couple ferredoxin oxidation or other kinds of cytoplasmic redox reactions with ion-translocation. \(^{21,24,42,43}\) and a V-type ATPase could be involved in ATP synthesis from the ion potential generated from both these complexes. Anaerobic Saganbacteria that also possess a Rnf complex and a V-type ATPase may be able to use this mechanism for energy conservation as well. However, the Rnf complex is reversible (like the Ech hydrogenase) and may play a role in substrate transport instead of ATP synthesis. \(^{24}\)

Two groups of closely related Saganbacteria have group 4 hydrogenases with unknown electron donors that may also be involved in the creation of a membrane potential. Saganbacteria like Rilefamargulisbacteria, could interconvert energy between an electrical gradient and an ionic gradient via group 4 NiFe hydrogenases. Given that the membrane-bound NiFe hydrogenases in Saganbacteria RX5A and RX6A have antiporter-like subunits, we suspect that they may function like Mbh hydrogenases in other organisms. For instance, in Thermococcus onnurineus, an Mbh hydrogenase produces H\(_2\) from the oxidation of reduced ferredoxin, creating a H\(^+\) potential across the membrane that is converted to a secondary Na\(^+\) gradient by the antipporter subunits. \(^{41,45}\) The energy stored in electrochemical gradients can be used by ATPases to synthesize ATP.

The possible FDH-Ehr complex in anaerobic Rilefamargulisbacteria and Saganbacteria is unusual, and may be involved in oxidative phosphorylation. In this scenario, the FDH-Ehr complex would serve as a simplified electron transport chain. We suspect that in Rilefamargulisbacteria and Saganbacteria the Nuo-like electron transfer subunits encoded next to the catalytic subunit of formate dehydrogenase (FdHA) are possibly part of a multimeric formate dehydrogenase (FDH). The formate dehydrogenase together with the hydrogenase-like subunits in the Ehr may pass electrons down to the transmembrane subunits, where an electron carrier such as quinone could be reduced and ion translocation would take place. For other Ehr complexes (e.g., Mbx), it was suggested that the distance between the formate dehydrogenase and membrane-bound subunits is important for the creation of an electrochemical potential across the membrane. \(^{46}\)

Experimental testing is needed to determine whether these Ehr complexes are part of a complex with formate dehydrogenase or another oxidoreductase, are expressed, assembled, and functional. The existence of an FDH-Ehr complex is even more questionable in Melainabacteria BJ4A, because the genes encoding the electron
transfer subunits in the vicinity of *fdhA* were not identified. Such complexes were not found in previously described Melainabacteria either. Regardless of the true function of the putative FDH-Ehr complexes, it is interesting to discover the variety of other membrane-bound energy-conserving complexes in lineages sibling to Cyanobacteria. Membrane-bound group 4 NiFe hydrogenases and related complexes share a common ancestor with Nuo, in which the catalytic subunits of NiFe hydrogenases may have later become the electron transport subunits of complex I. Nuo couples NADH oxidation with H⁺ or ion translocation and is a key component of respiratory and photosynthetic electron transport chains in other organisms.

Notable in Rillemargulisbacteria in particular (but present in other groups) are a variety of potential electron bifurcating complexes, including Nfn. Although not described as such, cytoplasmic NiFe hydrogenases (groups 3c) and cytoplasmic bidirectional FeFe hydrogenases could be involved in electron bifurcation as well. Electron bifurcating complexes are common in fermentative organisms, and those found in Rillemargulisbacteria probably serve as the primary mechanism to balance redox carriers in these organisms.

From the perspective of metabolic evolution, it is interesting to consider the origin of high electrode potential electron transport chains and their components. In this study, we were particularly interested in the electron entry (complex I) and exit points (terminal oxidases). Within the Saganbacteria and Melainabacteria, Saganbacteria HO1A and LO2A and Melainabacteria LO5B, HO7A and BJ4A are the only organisms that possess some sort of an electron transport chain. Given the prediction that the common ancestor of all groups was a fermentative anaerobe, we suspect that the ETC and ability to use O₂ was a later acquired trait.

Saganbacteria HO1A and LO2A and most Melainabacteria in this study (except AS1A and AS2A) have in common with Cyanobacteria the lack of the NADH module of NADH dehydrogenase (complex I). Respiration in the thylakoid membrane of Cyanobacteria is mostly initiated from succinate rather than NADH. However, how succinate dehydrogenase (SDH; complex II) works in *Synechocystis* sp. is not completely understood, in part due to the lack of two (out of four) subunits. This is also the case in Melainabacteria LO5B, HO7A and BJ4A. The complex I found in Saganbacteria HO1A and LO2A and Melainabacteria may be involved in proton translocation independent of NADH oxidation, and thus it may play a role similar to that of membrane-bound NiFe hydrogenases. Remarkably, Saganbacteria HO1A and LO2A also carry a putative novel type of heme-copper oxygen reductase (quinol oxidase), which could act as the terminal reductase in an electron transport chain.

Only marinamargulisbacteria is predicted to have an electron transport chain that includes an alternative complex III (ACIII) and a type A heme-copper oxygen reductase (cytochrome oxidase adapted to high O₂ levels), which may act as the intermediary between complex I and the terminal oxidase. Marinamargulisbacteria UBA6595 and Sericytromatia CBMW_127 are the only other related bacteria known to have an ACIII and a type A heme-copper oxygen reductase. Phylogenetic analyses indicate that Marinamargulisbacteria like Sericytromatia acquired this complex by lateral gene transfer. Given that Marinamargulisbacteria AA1A was found in the ocean, where dissolved O₂ may be available, it must have acquired the necessary machinery (i.e., complex I, menaquinone, ACIII, cytochrome oxidase) to take advantage of O₂ as a terminal electron acceptor. Similarly, fermentation and H₂ metabolism may be less relevant in the water column, and this clade may have lost the ancestral traits that were advantageous in other redox conditions.

The type C heme-copper oxygen reductase (complex IV) found in Melainabacteria HO7A and BJ4A may be adapted to low O₂ levels, as expected for microaerophilic or anoxic environments such as the subsurface and the human gut. Some aerobic Melainabacteria studied here, as well as *Obscuribacter phosphatis*, have a fusion involving complex III and complex IV. However, we annotated the proteins related to complex III in these Melainabacteria as cytochrome b₉f subunits. If correct, this is important because complex b₉f has only been found to play the role of complex III in Cyanobacteria. Thus, cytochrome b₉f complex may have been present in the common ancestor of all of these lineages.

Notably, Melainabacteria BJ4A could have a branched electron transport chain, with one branch leading to a cytochrome d ubiquinol oxidoreductase and the other leading to the type C heme-copper oxygen reductase. When two O₂ reductases are present in other lineages they tend to have different O₂ affinities. For instance, Cyanobacteria have a type A heme-copper oxygen reductase in the thylakoid membrane and a quinol cytochrome bd oxidase in the cytoplasmic membrane (in addition to a cytochrome bd oxidase). Sericytromatia also have two types of heme-copper oxygen reductases, type A and type C, that may also differ in their affinity for O₂.

Based on the absence of genes for CO₂ fixation and photosynthetic machinery in Melainabacteria and Sericytromatia, the lineages most closely related to Cyanobacteria, it was suggested that these capacities arose after their divergence. However, the alternative possibility is that they were a characteristic of their common ancestor, but lost in Melainabacteria and Sericytromatia. The current analyses support the former conclusion, given the essentially complete lack of genes that may be involved in carbon fixation and photosynthesis in two additional lineages sibling to Melainabacteria, Sericytromatia, and Cyanobacteria.

There is considerable interest in the metabolism of lineages sibling to the Cyanobacteria. Genomes from these lineages may provide clues to the origins of complexes that could have evolved to enable aerobic (and other types of) respiration via an electron transport chain. Based on the analyses presented here, we suggest that H₂ was central to the overall metabolism, and hydrogenases and the Rnf complex played central roles in proton translocation for energy generation, with redox carrier balance reliant upon electron bifurcation complexes.

**Methods**

**Samples collection, DNA extraction, and sequencing.** Publicly available genomes in this study were recovered from several sources (Supplementary Data 1). Margulisbacteria GW1B-GW1D, Saganbacteria HO1A, HO1B, LO2A, LO2B, HO2C, RX3A-RX8C, Melainabacteria HO7A, HO7A, LO5A, LO5B, RX6A-RX6C, GW8A, and GW9A originated from an alluvial aquifer in Rifle, CO, USA (groundwater samples; see Supplementary References). For sampling, DNA processing, sequencing, information, metagenome assembly, genome binning and curation of publicly available and newly generated genomes see Supplementary Methods.

**Data processing, assembly, binning, and curation of newly generated MAGs.** For description of reads trimming, assembly algorithm, binning methods and genome curation see Supplementary Methods.

**Completeness estimation.** Genome completeness was evaluated based on a set of 51 single copy genes previously used. Genomes from metagenomes and single-cell genomes were categorized according to the minimum information about a metagenome-assembled genome (MIMAG) and a single amplified genome (MISAG) of bacteria and archaea.

**Functional annotations.** Predicted genes were annotated using Prodigal and similarity searches were conducted using BLAST against UniProtKB and UniRef100, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) and uploaded to ggKbase (http://www.ggkbase.berkeley.edu). Additionally, the gene
products were scanned with hmmscan\(^{26}\) using an in house HMMs database representative of KEGG orthologous groups. More targeted HMMs were also used to confirm the presence of genes encoding the Pangenome. The improved HMMs were then used to identify NiFeDHk from all genomes in IMG/M using hmmssearch (hmmer 3.1b2, http://hmmer.org/). Significant hits for NiFeDHk were extracted from genomes which encoded all three genes. Sequences were aligned with mafft\(^{65}\), de-replicated by clustering with cd-hit\(^{59}\) at a similarity cutoff of 90%, and HMMs were built using hmmbuild (hmmer 3.1b2, http://hmmer.org/). Protein hits were extracted and de-replicated based on clustering of NifK at 90% sequence similarity with MCL\(^{62}\). NiFeDHk of cluster medoids were then aligned with mafft\(^{66}\), and alignments trimmed with trimmomatic\(^{27}\) to remove positions with more than 90% gaps. A phylogenetic tree was then constructed using a concatenated alignment of NiFeDHK with IQ-tree with 100 bootstrap support and the numbers of bootstraps were indicated.

**Catalytic subunit of dimethyl sulfoxide reductase superfamily protein and catalytic subunit I of heme–copper oxygen reductases phylogenetic trees.** Each individual protein data set was aligned using Muscle version 3.8.31\(^{75,76}\) and then manually curated to remove end gaps. Phylogenies were conducted using RAxML-HPC-BlackBox\(^{77}\) as implemented on the CIPRES server\(^{78}\) under the PROTGAMMA JTT evolutionary model and with the number of bootstrap support counts determined.

**HCO A-family oxygen reductase protein.** Homologs of CoaXa in Marinarchaea and Saganbacteria genomes were identified from HMM searches. Other homologs were gathered from Shih et al\(^{2}\). Sequences were aligned with MAFFT using the -maxiterate\(^{79}\). Phylogenetic analysis was performed using RAxML through the CIPRES Science Gateway\(^{77}\) under the LG model.

**Statement of Ethics.** The fecal samples obtained were part of a clinical phase/II study in rural Bangladesh entitled “Selenium and arsenic pharmacodynamics” (SEASP) run by Graham George (University of Saskatchewan) and the SEASP trial was approved by the University of Saskatchewan Research Ethics Board (14-284) and the Bangladesh Medical Research Council (940,BMRC/NREC/2010-2013/291). Additional ethics approval was also obtained by UCL (7591/001). The study complied with all the relevant ethical regulations. Informed consent was obtained from all human participants.

**Data availability**

New and published genomes included in this study and corresponding gene annotations can be accessed at https://ggkbbase.berkeley.edu/Margarulis_Sagan_Melanotaeniarchaeota/organisms (ggkbase is a ‘live’ site, genomes may be updated after publication). DNA sequences (new genomes and raw sequence reads) have been deposited in the NCBI Bioproject Database (accession codes: PRJNA167727, PRJNA451230, PRJNA471730, PRJNA471718). Further details are provided in Supplementary Data 1, including NCBI Genbank accession numbers for individual genomes. A reporting summary for this Article as a Supplementary Information file.

Received: 25 May 2018 Accepted: 18 December 2018
Published online: 28 January 2019

**References**

1. Luo, G. et al. Rapid oxygenation of Earth’s atmosphere 2.33 billion years ago. Sci. Adv. 2, e1600134 (2016).
2. Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J. & Fischer, W. W. Crown group oxyphotobacteria postdate the rise of oxygen. Geobiology 15, 19–29 (2017).
3. Stal, L. J. & Morel, F. Reformation in cyanobacteria. FEMS Microbiol. Rev. 21, 179–211 (1997).
4. Refojo, P. N., Teixeira, M. & Pereira, M. M. The alternative complex III: properties and possible mechanisms for electron transfer and energy conservation. Biochim. Biophys. Acta 1817, 1852–1859 (2012).
5. Dit-Renz, S. C. et al. The human gut and groundwater harbor nonphotosynthetic bacteria belonging to a new candidate phylum sibling to cyanobacteria. Elife 2, https://doi.org/10.7554/eLife.01102 (2013).
6. Søo, R. M. et al. An expanded genomic representation of the phylum Cyanobacteria. Genome Biol. Evol. 6, 1031–1045 (2014).
7. Søo, R. M., Hemp, J., Parks, D. H., Fischer, W. W. & Hugenholtz, P. On the origins of oxygenic photosynthesis and aerobic respiration in cyanobacteria. Nature 435, 1436–1440 (2005).
8. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7, https://doi.org/10.1038/ncomms13219 (2016).
Acknowledgements

Laura Hug and Christopher Brown provided assistance with analyses, Shufei Lei and Kate Lane assisted with bioinformatics and data management, Kelly Wrighton, Kim Handley, and Kenneth Hurst Williams provided samples. Sallie Chisholm, Paul Berube, and Steven Biller are acknowledged for their help securing the marine samples. We thank the staff of Bigelow Laboratory Single-Cell Genomics Center for the generation of single-cell data. Teruki Iwatsuki, Kazuki Hayashida, Toshihiro Kato, and Mitsuru Kubota assisted with groundwater sampling at Mizunami Underground Research Laboratory, Japan Atomic Energy Agency (JAEA). Thanks to Chris Greening for feedback regarding FDM-Ehr complexes on the bioRxiv version of this manuscript, and to Denis Baurain and an anonymous reviewer for all the useful comments on the manuscript. The research was supported by the Department of Energy (DOE), Office of Science and Office of Biological and Environmental Research (Lawrence Berkeley National Lab; Operated by the University of California, Berkeley). DNA sequencing for the Rifle samples was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under Contract No. DE-AC02-05CH11231. Marine single amplified genomes were generated and sequenced with the support of NSF grants DEB-1441717 and OCE-1335810, and Simons Foundation grant 510023 (to R.S.). Fecal samples were collected from patients in the clinical Phase I/II SEASP trial in Bangladesh that was jointly led by Graham George and Ingrid Pickering (University of Saskatchewan), with the assistance of the SEASP team https://clinicaltrials.gov/ct2/show/NCT02377635, and funded by the Canadian Federal Government, through Grand Challenges Canada, Stars in Global Health and by the Global Institute for Water Security. The study was funded by the Canadian Federal Government, through a program entitled Grand Challenges Canada, Stars in Global Health, with additional funds from the Global Institute for Water Security at the University of Saskatchewan.

Author contributions

P.B.M.C., I.S., F.S., B.C.T., M.R.O., and J.F.B. reconstructed and curated the genomes; P.B.M.C. conducted the majority of the metabolic analyses; F.S., C.J.C., R.K., D.B., P.S., K.A., and J.F.B. provided input to analyses; F.S., C.J.C., P.S., and P.B.M.C. generated phylogenetic trees; Y.A., E.D.B., and R.S. acquired samples; Y.A., J.M.S., E.D.B., and R.S. acquired new sequence information; and E.D.B., R.S., and T.W. contributed single amplified genomes (SAGs). P.B.M.C. and J.F.B. wrote the manuscript with input from F.S. and T.W., as well as C.J.C., P.S., and R.K.; All authors reviewed the results and approved the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-08246-y.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019