The ISME Journal (2016) 10, 761–777
© 2016 International Society for Microbial Ecology All rights reserved 1751-7362/16
www.nature.com/ismej

ORIGINAL ARTICLE

Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival

Chris Greening[1,2], Ambarish Biswas[1], Carlo R Carere[3,4], Colin J Jackson[5], Matthew C Taylor[2], Matthew B Stott[3], Gregory M Cook[1,6] and Sergio E Morales[1]

[1]*Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand;* [2]*The Commonwealth Scientific and Industrial Research Organisation, Land and Water Flagship, Acton, Australian Capital Territory, Australia;* [3]*GNS Science, Wairakei Research Centre, Taupō, New Zealand;* [4]*Scion, Te Papa Tipu Innovation Park, Rotorua, New Zealand;* [5]*Australian National University, Research School of Chemistry, Acton, Australian Capital Territory, Australia and* [6]*University of Auckland, Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand*

Recent physiological and ecological studies have challenged the long-held belief that microbial metabolism of molecular hydrogen (H_2) is a niche process. To gain a broader insight into the importance of microbial H_2 metabolism, we comprehensively surveyed the genomic and metagenomic distribution of hydrogenases, the reversible enzymes that catalyse the oxidation and evolution of H_2. The protein sequences of 3286 non-redundant putative hydrogenases were curated from publicly available databases. These metalloenzymes were classified into multiple groups based on (1) amino acid sequence phylogeny, (2) metal-binding motifs, (3) predicted genetic organisation and (4) reported biochemical characteristics. Four groups (22 subgroups) of [NiFe]-hydrogenase, three groups (6 subtypes) of [FeFe]-hydrogenases and a small group of [Fe]-hydrogenases were identified. We predict that this hydrogenase diversity supports H_2-based respiration, fermentation and carbon fixation processes in both oxic and anoxic environments, in addition to various H_2-sensing, electron-bifurcation and energy-conversion mechanisms. Hydrogenase-encoding genes were identified in 51 bacterial and archaeal phyla, suggesting strong pressure for both vertical and lateral acquisition. Furthermore, hydrogenase genes could be recovered from diverse terrestrial, aquatic and host-associated metagenomes in varying proportions, indicating a broad ecological distribution and utilisation. Oxygen content (pO_2) appears to be a central factor driving the phylum- and ecosystem-level distribution of these genes. In addition to compounding evidence that H_2 was the first electron donor for life, our analysis suggests that the great diversification of hydrogenases has enabled H_2 metabolism to sustain the growth or survival of microorganisms in a wide range of ecosystems to the present day. This work also provides a comprehensive expanded system for classifying hydrogenases and identifies new prospects for investigating H_2 metabolism.
The ISME Journal (2016) **10,** 761–777; doi:10.1038/ismej.2015.153; published online 25 September 2015

Introduction

Molecular hydrogen (H_2) has several physical properties desirable for biological systems, notably its redox potential ($E^{\circ\prime} = -0.42$ V) and diffusion coefficient (4×10^{-9} $m^2 s^{-1}$). Microorganisms are able to harness these properties by consuming and producing H_2 using specialised metalloenzymes called hydrogenases (Schwartz *et al.*, 2013). There are three phylogenetically unrelated classes of hydrogenase distinguishable based on the metal content of their H_2-binding sites: the [NiFe]-, [FeFe]- and [Fe]-hydrogenases (Volbeda *et al.*, 1995; Peters *et al.*, 1998; Shima *et al.*, 2008). H_2 oxidation by such enzymes yields low-potential electrons that are transduced through respiratory chains or used to fix inorganic carbon. In contrast, H_2 evolution efficiently dissipates excess reductant as a diffusible gas during microbial fermentation and photobiological processes (Schwartz *et al.*, 2013). Certain hydrogenases are also part of low-potential ion-translocating complexes that use protons as terminal electron acceptors (Buckel and Thauer, 2013).

Correspondence: C Greening, The Commonwealth Scientific and Industrial Organisation, Land and Water Flagship, Clunies Ross Street, Acton, Australian Capital Territory 2060, Australia.
E-mail: chris.greening@csiro.au
or S Morales, Department of Microbiology and Immunology, University of Otago, 720 Cumberland Street, North Dunedin, Dunedin 9054, New Zealand.
E-mail: sergio.morales@otago.ac.nz
Received 3 May 2015; revised 20 June 2015; accepted 20 July 2015; published online 25 September 2015

Since the discovery of microbial H_2 oxidation in the 1900s (Kaserer, 1906; Stephenson and Stickland, 1931), H_2 metabolism has been observed in multiple bacterial, archaeal and eukaryotic phyla. It is increasingly recognised that H_2 metabolism is important for a wide range of microorganisms: lithotrophs and phototrophs, respirers and fermenters and aerobes and anaerobes alike (Vignais and Billoud, 2007; Schwartz *et al.*, 2013; Peters *et al.*, 2014). Furthermore, it is widely hypothesised that H_2 was the primordial electron donor, suggesting early and sustained evolutionary importance (Lane *et al.*, 2010).

Several recent studies demonstrated that microbial H_2 metabolism is more widespread than previously reported. It was recently shown that some aerobic soil actinobacteria and acidobacteria persist by scavenging H_2 from the lower atmosphere (Constant *et al.*, 2010; Greening *et al.*, 2014, 2015a, b), overturning long-held beliefs that hydrogen metabolism is restricted to low O_2, high H_2 environments and highlighting the importance of H_2 for survival in addition to growth (Greening and Cook, 2014). Biochemists have simultaneously elucidated mechanisms dependent on reversed electron flow that enable certain hydrogenases to function in the presence of the oxygen (traditionally an inhibitor of their active sites) (Fritsch *et al.*, 2011; Shomura *et al.*, 2011; Horch *et al.*, 2015). In anaerobic systems, ultra-minimalistic hydrogenase-containing respiratory chains have been described that efficiently generate energy within oligotrophic environments (Kim *et al.*, 2010; Lim *et al.*, 2014). In parallel, the discovery of electron bifurcation has expanded our understanding of how energy is conserved in anaerobic processes such as cellulolytic fermentation, acetogenesis and methanogenesis (Schut and Adams, 2009; Kaster *et al.*, 2011; Buckel and Thauer, 2013; Schuchmann and Muller, 2014). Other themes, including H_2 sensing within anaerobes (Zheng *et al.*, 2014) and H_2 fermentation in aerobes (Berney *et al.*, 2014), are emerging.

Despite this progress, there remains much to be discovered about microbial H_2 metabolism on both the microscopic and macroscopic levels. Most studies on microbial H_2 metabolism focus on only a few branches of the hydrogenase phylogenetic tree and a small subset of organisms within the universal tree of life. Physiological and biochemical characterisations have focussed on model organisms from within five phyla, *Proteobacteria*, *Firmicutes*, *Cyanobacteria*, *Euryarchaeota* and *Chlorophyta* (Schwartz *et al.*, 2013; Lubitz *et al.*, 2014). Furthermore, detailed biochemical information and atomic-resolution structures are available for only a subset of hydrogenases (Volbeda *et al.*, 1995; Peters *et al.*, 1998; Shima *et al.*, 2008; Fritsch *et al.*, 2011; Mills *et al.*, 2013). Although the contribution of H_2 metabolism to total ecosystem processes is recognised in some environments (for example, anoxic sediments, animal guts and hydrothermal vents; Vignais and Billoud, 2007; Schwartz *et al.*, 2013), the role of hydrogenases in general soil and aquatic ecosystems remains largely unresolved (Barz *et al.*, 2010; Constant *et al.*, 2011; Beimgraben *et al.*, 2014; Greening *et al.*, 2015b). Consequently, the influence of H_2 evolution and consumption on community structuring and global biogeochemical cycling requires further investigation (Schwartz *et al.*, 2013; Greening *et al.*, 2015b).

Hydrogenase gene surveys are vital for understanding microbial H_2 metabolism at the global scale. Current knowledge on the evolution and diversity of hydrogenases relies heavily on the progressive surveys conducted by Wu and Vignais (Wu and Mandrand, 1993; Vignais *et al.*, 2001; Vignais and Billoud, 2007); these studies revealed that the primary sequences and subunit architectures of [NiFe]- and [FeFe]-hydrogenases have diversified to enable them to adopt a wide range of physiological roles (whereas the [Fe]-hydrogenase is constrained to a single function). In the eight years following these studies (Vignais and Billoud, 2007), the emergence of sequencing technologies has resulted in the rapid expansion of genome and metagenome sequence data. Genomes are now available for a far greater range of organisms, spanning model laboratory specimens, representatives of dominant environmental phyla, and poorly described 'Microbial Dark Matter' (Wu *et al.*, 2009; Rinke *et al.*, 2013). Furthermore, metagenomes enable the metabolic capability of entire communities to be described *in silico* (Tringe *et al.*, 2005; Morales and Holben, 2011; Wrighton *et al.*, 2012). In this work, we used publicly available genome and metagenome resources to comprehensively analyse the distribution of hydrogenases. Our findings suggest that H_2 metabolism is more diverse and widespread on both the taxonomic and community levels than previously reported.

Materials and methods

Hydrogenase sequence retrieval

Amino acid sequences of all non-redundant putative hydrogenase catalytic subunits represented in the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) (Pruitt *et al.*, 2007) and Joint Genome Institute (JGI) Microbial Dark Matter (MDM) (Rinke *et al.*, 2013) databases were retrieved by Protein BLAST (Altschul *et al.*, 1990) during August 2014. The retrieved sequences were verified as hydrogenase encoding through screening for the presence of conserved cysteine residues required to ligate H_2-binding metal centres (L1 and L2 motifs for [NiFe]-hydrogenases (Vignais and Billoud, 2007); P1, P2 and P3 motifs for [FeFe]-hydrogenases (Vignais and Billoud, 2007); and Cys176 in [Fe]-hydrogenases (Shima *et al.*, 2008)). The analysis omitted protein families

homologous to [NiFe]-hydrogenases (Ehr/Mbx, NuoD), [FeFe]-hydrogenases (Narf/Nar1p) and [Fe]-hydrogenases (HmdII) that appear to lack the capacity to metabolise H_2. Our analysis did not include nitrogenases, alkaline phosphatases, formate dehydrogenases and carbon monoxide dehydrogenases that have been shown to catalyse side-reactions resulting in H_2 oxidation or evolution (Schwartz *et al*., 2013).

Hydrogenase classification and analysis
Protein sequences encoding the catalytic subunits ([NiFe]-hydrogenases, [Fe]-hydrogenases) or catalytic domains ([FeFe]-hydrogenases) of hydrogenases were aligned using the ClustalW (Larkin *et al*., 2007) and MUSCLE (Edgar, 2004) algorithms. Evolutionary relationships were analysed using neighbour-joining phylogenetic trees (Saitou and Nei, 1987) constructed with MEGA6 (Tamura *et al*., 2013). All trees were bootstrapped using 500 replicates and were rooted with ancestral sequences where available. The robustness of the analysis was confirmed by varying the number of ingroup sequences tested and the nature of the outgroup sequence used. The Microbial Genomic Context Viewer (MGcV) was used to compare genome regions encoding homologous hydrogenases (Overmars *et al*., 2013). Domains were predicted by searching the Conserved Domain Database (CDD) (Marchler-Bauer *et al*., 2011) and using multiple sequence alignments to identify signature conserved residues. WebLogo (Crooks *et al*., 2004) was used to visualise conserved metal-binding motifs forming the active sites and redox centres. Using a combination of the information derived from these methods, the [NiFe]- and [FeFe]-hydrogenases were further divided into phylogenetically distinct groups and subgroups.

Metagenome analysis
Metagenome sequence libraries derived from 10 ecosystems including soil (farmland, forest, permafrost, bog), gut (termite, human) and water (fresh water, hot spring, coastal upwelling, deep ocean) environments were identified from publicly available databases. For each ecosystem, two libraries were selected. All selected libraries were sequenced with paired-end reads on an Illumina (San Diego, CA, USA) platform and represented read sizes of between 201 and 280 nucleotides. BLAST (Altschul *et al*., 1990) analyses were performed using a local BLAST database containing the protein sequences of the catalytic subunit ([NiFe]-hydrogenases, [Fe]-hydrogenases) or catalytic domain ([FeFe]-hydrogenases) of all sequenced hydrogenases (curated as described above). Low complexity regions for all reference sequences were masked using the SEG algorithm (Wootton and Federhen, 1996) of BLAST+ (Camacho *et al*., 2008) and a reference BLAST database was created. All metagenome libraries were randomly subsampled to an equal depth (1 million reads) and read length >201 nucleotides before analyses. A translated BLAST screening of all subsamples was performed using blastx (word size 3 and *e*-value 10). To minimise false positives, hits within the initial screen were sieved by removing any result with a minimum percentage identity 60% and minimum query coverage 40 amino acids. Identified reads for each class are recorded as relative percentage abundance.

Results

An expanded hydrogenase classification scheme predictive of biological function
The first aim of this work was to identify and classify all putative hydrogenases represented in public databases. Initially, we retrieved non-redundant sequences encoding the catalytic subunits of all hydrogenases in the NCBI database and verified that 3286 of them contained the sufficient residues required to bind their metal centres (Supplementary Tables S1 and S2). In order to develop the classification scheme (Table 1), we correlated the phylogenetic clustering of the hydrogenases with functional information and predictors. For all hydrogenases, we analysed: (1) primary phylogeny to determine their evolutionary relationships (Figure 1), (2) metal- and cofactor-binding motifs to predict redox centres (Table 2), (3) genetic organisation to identify probable partner proteins (Figure 2) and (4) previous literature reports to probe biochemical characteristics and physiological roles (Table 1). Integrating this information, we were able to classify hydrogenases into multiple groups and subgroups/subtypes likely to have distinct cellular functions.

All hydrogenases could be classified into eight previously described major lineages (Vignais and Billoud, 2007; Calusinska *et al*., 2010): groups 1 to 4 [NiFe]-hydrogenases, groups A to C [FeFe]-hydrogenases and [Fe]-hydrogenases. However, pre-existing classification schemes did not sufficiently reflect the variety in the functions of the experimentally studied hydrogenases within most of these groups. For example, existing schemes do not account for the great heterogeneity in primary sequence phylogeny, genetic organisation and physiological roles of the group 1 and group 4 [NiFe]-hydrogenases (Vignais and Billoud, 2007), as well as the group A [FeFe]-hydrogenases (Calusinska *et al*., 2010). In addition, the group 2 [NiFe]-hydrogenases of recently sequenced *Aquificae* and methanotrophs form distinct lineages from the presently recognised 2a and 2b subgroups (Vignais and Billoud, 2007). We therefore expanded the [NiFe] enzymes into 22 functionally distinct subgroups (groups 1a to 1h, 2a to 2d, 3a to 3d and 4a to 4f) and the group A [FeFe]-hydrogenases into four subtypes (groups A1 to A4).

Table 1 New scheme for the classification of hydrogenases

Group	*Proposed function*	*References*	*PDB*
Group 1: membrane-bound H_2-uptake [NiFe]-hydrogenases			
Group 1a: ancestral	Liberates electrons for sulphate, metal, organohalide and methanogenic heterodisulphide respiration. Includes [NiFeSe] variants.	*Desulfovibrio vulgaris* (Marques *et al*., 2010) *Desulfomicrobium baculatum* (Garcin *et al*., 1999) *Methanosarcina mazei* (Deppenmeier and Blaut, 1995)	2WPN 1CC1
Group 1b: prototypical	Liberates electrons for sulphate, fumarate and nitrate respiration.	*Desulfovibrio gigas* (Volbeda *et al*., 1995) *Wolinella succinogenes* (Gross *et al*., 1998) *Helicobacter pylori* (Olson and Maier, 2002)	1FRV
Group 1c: Hyb type	Liberates electrons primarily for fumarate respiration. Possibly bidirectional.	*Escherichia coli* (Lukey *et al*., 2010) *Salmonella enterica* (Maier *et al*., 2013)	
Group 1d: oxygen tolerant	Electron input for aerobic respiration and oxygen-tolerant anaerobic respiration.	*Ralstonia eutropha* (Fritsch *et al*., 2011) *Escherichia coli* (Volbeda *et al*., 2013) *Aquifex aeolicus* (Brugna-Guiral *et al*., 2003)	3RGW 4GD3
Group 1e: Isp type	Liberates electrons primarily for sulphur respiration. Possibly bidirectional.	*Allochromatium vinosum* (Ogata *et al*., 2010) *Aquifex aeolicus* (Brugna-Guiral *et al*., 2003) *Thiocapsa roseopersicina* (Tengölics *et al*., 2014)	3MYR
Group 1f: oxygen protecting	Unresolved. May liberate electrons to reduce reactive oxygen species.	*Geobacter sulfurreducens* (Tremblay and Lovley, 2012) *Frankia* sp. (Leul *et al*., 2007)	
Group 1g: Crenarchaeota type	Unresolved. May liberate electrons primarily for sulphur respiration.	*Pyrodictium brockii** (Pihl *et al*., 1989) *Acidianus ambivalens** (Laska *et al*., 2003)	
Group 1h/5: Actinobacteria type	Scavenges electrons from tropospheric H_2 to sustain aerobic respiration during starvation.	*Ralstonia eutropha* (Schäfer *et al*., 2013) *Mycobacterium smegmatis* (Greening *et al*., 2014) *Streptomyces avermitilis* (Constant *et al*., 2010)	5AA5
Group 2: Cytosolic H_2-uptake [NiFe]-hydrogenases			
Group 2a: Cyanobacteria type	Electron input for aerobic respiration and recycling H_2 produced by cellular processes (for example, nitrogenase, fermentation).	*Anabaena* sp. (Houchins and Burris, 1981) *Mycobacterium smegmatis* (Greening *et al*., 2014) *Nitrospira moscoviensis** (Koch *et al*., 2014)	
Group 2b: HK linked	Senses H_2 and activates two-component cascade controlling hydrogenase expression.	*Ralstonia eutropha* (Lenz and Friedrich, 1998) *Rhodobacter capsulatus* (Vignais *et al*., 2005)	
Group 2c: DGC linked (putative)	Unknown. Predicted to sense H_2 and induce cyclic di-GMP production.	Uncharacterised.	
Group 2d: Aquificae type	Unknown. May generate reductant for carbon fixation or have a regulatory role.	*Aquifex aeolicus* (Brugna-Guiral *et al*., 2003)	
Group 3: Cytosolic bidirectional [NiFe]-hydrogenases			
Group 3a: F_{420} coupled	Directly couples oxidation of H_2 to reduction of F_{420} during methanogenesis. Reverse reaction may also occur. Includes [NiFeSe] variants.	*Methanothermobacter marburgensis* (Mills *et al*., 2013) *Methanosarcina barkeri* (Kulkarni *et al*., 2009)	4OMF
Group 3b: NADP coupled	Directly couples oxidation of NADPH to evolution of H_2. May be reversible. Some complexes are proposed to have sulfhydrogenase activity.	*Pyrococcus furiosus* (Ma *et al*., 1993) *Thermococcus kodakarensis* (Kanai *et al*., 2011) *Mycobacterium smegmatis* (Berney *et al*., 2014)	
Group 3c: HDR linked	Bifurcates electrons from H_2 to heterodisulphide and ferredoxin in methanogens without cytochromes.	*Methanothermobacter marburgensis* (Kaster *et al*., 2011)	
Group 3d: NAD coupled	Directly interconverts electrons between H_2 and NAD depending on redox state.	*Anabaena* sp. (Houchins and Burris, 1981) *Ralstonia eutropha* (Burgdorf *et al*., 2005) *Thiocapsa roseopersicina* (Rákhely *et al*., 2004)	
Group 4: Membrane-bound H_2-evolving [NiFe]-hydrogenases			
Group 4a: formate hydrogenlyases	Couples oxidation of formate to fermentative evolution of H_2. Hyf-type complexes may translocate protons via antiporter modules.	*Escherichia coli* (McDowall *et al*., 2014; Andrews *et al*., 1997) *Salmonella enterica* (Sawers *et al*., 1986)	
Group 4b: Mrp linked	Couples oxidation of formate or carbon monoxide to proton reduction. Generates sodium-motive force via Mrp antiporter modules.	*Pyrococcus furiosus* (Sapra *et al*., 2003) *Thermococcus onnurineus* (Lim *et al*., 2014) *Thermococcus kodakarensis* (Kanai *et al*., 2011)	
Group 4c: CODH linked	Forms complex with carbon monoxide dehydrogenase to anaerobically respire CO using protons as terminal electron acceptors.	*Carboxydothermus hydrogenoformans* (Soboh *et al*., 2002) *Rhodospirillum rubrum* (Fox *et al*., 1996)	
Group 4d: Eha/Ehb type	Multimeric complexes that couple H_2 oxidation to ferredoxin reduction for anaplerotic (Eha) and anabolic (Ehb) purposes. H^+/Na^+ driven.	*Methanococcus maripaludis* (Lie *et al*., 2012; Porat *et al*., 2006)	

Table 1 (Continued)

Group	*Proposed function*	*References*	*PDB*
Group 4e: Ech type	Couples ferredoxin oxidation to H_2 evolution. This process is physiologically reversible via H^+/Na^+ translocation.	*Methanosarcina barkeri* (Meuer and Kuettner, 2002) *Desulfovibrio gigas* (Morais-Silva *et al*., 2013) *Thermoanaerobacter tengcongensis* (Soboh *et al*., 2004)	
Group 4f: Ehf type (putative)	Unknown. May couple oxidation of a one-carbon compound to proton reduction concurrent with proton translocation. Related to Ehr complexes.	Uncharacterised.	
[FeFe]-hydrogenases			
Group A1: prototypical	Couples oxidation of ferredoxin to fermentative or photobiological evolution of H_2.	*Clostridium pasteurianum* (Peters *et al*., 1998) *Desulfovibrio desulfuricans* (Nicolet *et al*., 1999) *Chlamydomonas reinhardtii* (Happe and Naber, 1993)	1FEH 1HFE 3LX4
Group A2: glutamate synthase linked (putative)	Unknown. Predicted to transfer electrons from H_2 to NAD, generating reducing equivalents for glutamate synthase.	Uncharacterised.	
Group A3: bifurcating	Reversibly bifurcates electrons from H_2 to ferredoxin and NAD in anaerobic bacteria.	*Thermotoga maritima* (Schut and Adams, 2009) *Acetobacterium woodii* (Schuchmann and Müller, 2012) *Moorella thermoacetica* (Wang *et al*., 2013)	
Group A4: formate dehydrogenase linked	Couples formate oxidation to evolution of H_2. Some can also bifurcate electrons from H_2 to ferredoxin and NADP.	*Clostridium autoethanogenum* (Wang *et al*., 2013)	
Group B: ancestral (putative)	Unknown. May couple oxidation of ferredoxin to fermentative evolution of H_2.	Uncharacterised.	
Group C: sensory (putative)	Unknown. Predicted to sense hydrogen and induce cascades via co-transcribed regulatory elements, for example, kinases and phosphatases.	*Thermoanaerobacterium saccharolyticum* (Shaw *et al*., 2009) *Ruminococcus albus* (Zheng *et al*., 2014)	
[Fe]-hydrogenases			
HmdI: methenyl-H4MPT dehydrogenase	Couples oxidation of H_2 to reduction of 5,10-methenyltetrahydromethanopterin in methanogens. Physiologically-reversible and important during nickel limitation.	*Methanocaldococcus jannaschii* (Shima *et al*., 2008) *Methanothermobacter thermoautotrophicum* (Afting *et al*., 1998)	3DAG

This scheme correlates the phylogenetic clustering of the hydrogenases with their probable functions. The groups A, B and C [FeFe]-hydrogenases and several subgroups of the [NiFe]-hydrogenases (2a, 2b, 3a, 3b, 3c, 3d) were previously defined (Vignais *et al*., 2001; Vignais and Billoud, 2007; Calusinska *et al*., 2010). The group 1h [NiFe]-hydrogenases have also been defined as the group 5 [NiFe]-hydrogenases; however, this work shows they are they are descended from other group 1 [NiFe]-hydrogenases. References are provided to structural, biochemical and physiological characterisations of representative hydrogenases from each subgroup/subtype. Structural characterisations are listed first and Protein Database (PDB) structures are provided (see Lubitz *et al*., 2014 for a full list of solved hydrogenase structures). Asterisked organisms have yet to be sequenced, hence their [NiFe]-hydrogenases are not represented on the phylogenetic trees.

A description of the hydrogenase subgroups/subtypes defined here is provided in Table 1.

Functional diversity of hydrogenases is reflected in phylogenetic clustering, genetic organisation and metal-binding motifs

The hydrogenase classification scheme was developed primarily on the basis of amino acid sequence phylogeny (Figure 1). Our analysis shows that the [NiFe]-hydrogenases are the most diverse and widespread of the hydrogenases, and can be initially divided into H_2-uptake (groups 1 and 2), bidirectional (group 3) and H_2-evolving (group 4) clades. On the basis of phylogeny, these enzymes can be further divided into 22 subgroups (Figure 1; Supplementary Figures S1–S4) predicted to have distinct physiological roles described in Table 1. These subgroups are monophyletic, well populated (with more than 15 unique sequences) and statistically supported (with bootstrap values above 0.75), with the exceptions of certain group 4 clades (see Supplementary Figure S4 legend). On the basis of phylogeny, three [FeFe]-hydrogenase groups were defined: a main group represented by fermentative and bifurcating hydrogenases (group A), an ancestral group of unknown function (group B) and a group containing putative sensory hydrogenases (group C) (Figure 1). However, we were unable to subdivide these enzymes further by phylogeny alone, owing to poor bootstrapping of the group A subclades (Supplementary Figure S5) and lack of functional information on the group B and C enzymes.

The genetic organisation of the hydrogenases also serves as a reliable indicator of function between subgroups. As detailed in Figure 2, genes encoding hydrogenase structural components are proximal to those encoding diverse electron-transfer proteins, ion-translocating subunits, regulatory components, maturation factors, hypothetical proteins and partner

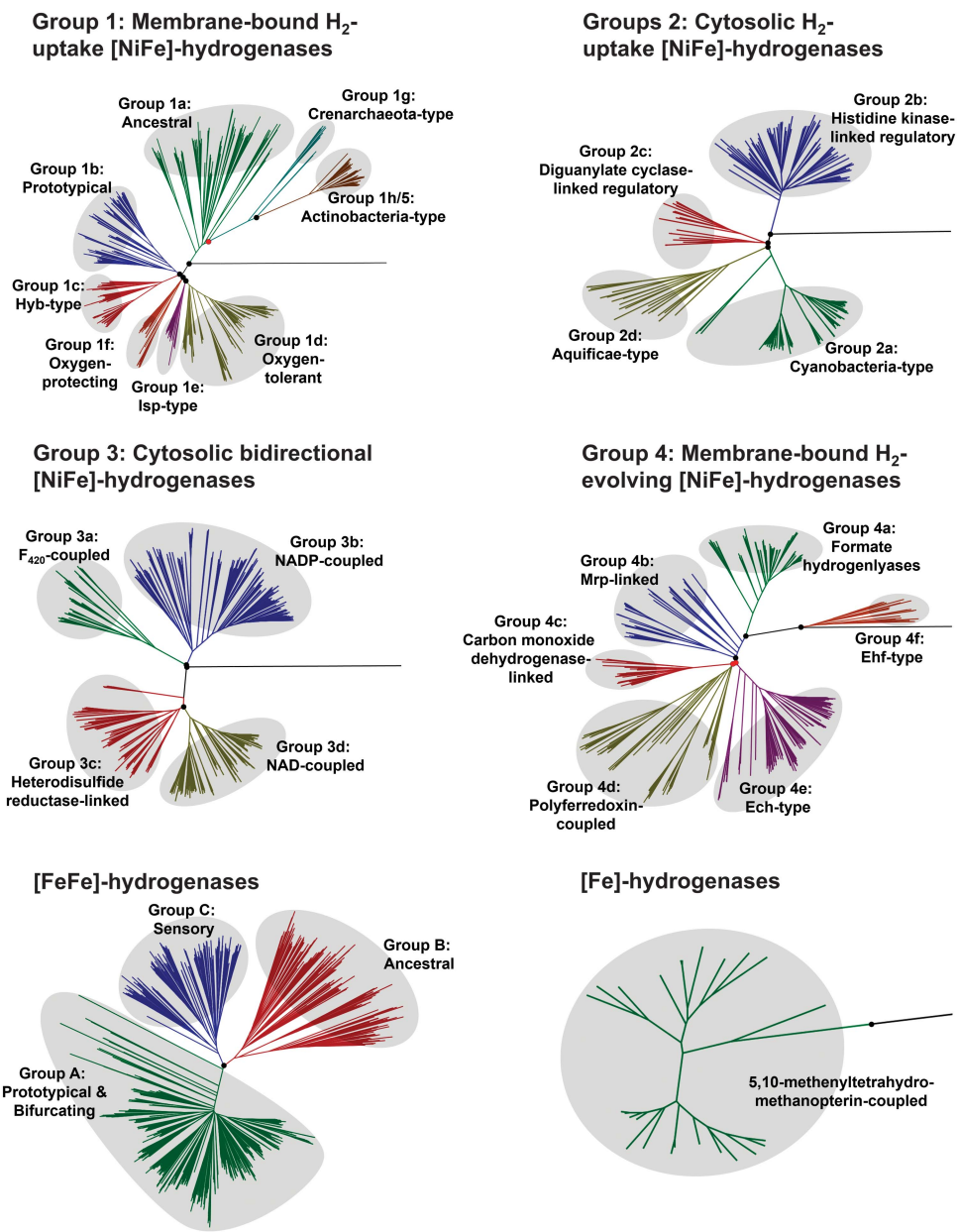


Figure 1 Classification and phylogeny of hydrogenases. These neighbour-joining skeleton trees show the phylogenetic relationships of all 3286 hydrogenases identified in this work. The trees are colour coded by [NiFe]-hydrogenase subgroup and [FeFe]-hydrogenase group. The nodes separating the major clades are encircled and coloured according to their bootstrap values, that is, black circles for well-supported nodes (bootstrap values >0.75) and red circles for unsupported nodes (bootstrap values <0.75). Group A [FeFe]-hydrogenases cannot be reliably subdivided phylogenetically and can only be classified into subtypes based on their genetic organisation. The expanded trees, including taxon names and bootstrap values, are shown in Supplementary Figures S1 to S6.

enzymes. Although many of these associations have been previously described (Vignais and Billoud, 2007; Schwartz *et al.*, 2013), novel findings included the association of surprising regulatory components (for example, diguanylate cyclases/phosphodiesterases) with putative group C [FeFe]-hydrogenases and group 2c [NiFe]-hydrogenases. Genome architecture is well conserved within subgroups of [NiFe]-hydrogenases, but varies extensively between enzymes of different functions (Figure 2). Genome architecture is also able to discriminate the sometimes poorly bootstrapped lineages of the group 4 enzymes (Supplementary Figure S4) and therefore is a valuable hydrogenase classification tool. Variations in the domain organisation and probable quaternary structure of the [FeFe]-hydrogenases were also apparent. On this basis, group A [FeFe]-hydrogenases can be subdivided into four functionally distinct subtypes: stand-alone enzymes (group A1) and those associated with putative glutamate synthases (group A2), NADH dehydrogenases (group A3) and formate dehydrogenases (group A4) (Figure 2).

Table 2 Metal centres of hydrogenases

[NiFe]	*L1 motif*	*L2 motif*	*[FeS]$_{proximal}$*	*[FeS]$_{medial}$*	*[FeS]$_{distal}$*	*Extension*
Group 1a	xxRICGVCPxxH	SFDPCxxCxxH*	4Cys[4Fe4S]	4Cys[4Fe4S]	3Cys1His/Asp[4Fe4S]	Y
Group 1b	xQRxCGVCTxxH	xxDPCxACxVH	4Cys[4Fe4S]/3Cys1Asn[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 1c	xQRICGVCTTVH	SFDPCxxCAVH	4Cys[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 1d	xxRICGVCTxxH	SFDPCLACxxH	6Cys[4Fe3S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 1e	xxRICGVCTxVH	SFDPCxACxxH	4Cys[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 1f	xQRxCGVCTxVH	SFDPCxACxVH	4Cys[4Fe4S]/3Cys1Asp/Asn[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 1g	xSRxCGVCGxxH	SFDPCxxCxVH	3Cys1Asp/Asn[4Fe4S]	3Cys[3Fe4S]	3Cys1His/Arg[4Fe4S]	Y
Group 1h	TSRICGICGDNH	SFDPCLPCGVH	3Cys1Asp[4Fe4S]	4Cys[4Fe4S]	3Cys1His[4Fe4S]	Y
Group 2a	xxRICGICGxxH	SFDxCLVCTVH	3Cys1Asn[4Fe4S]	3Cys[3Fe4S]	3Cys1Glu[4Fe4S]	Y
Group 2b	xPRICGICSxSQ	SFDPCMVCTVH	4Cys[4Fe4S]	4Cys[4Fe4S]	3Cys1His[4Fe4S]	N
Group 2c	xxRxCGICxxxH	SxDPCxxCTVH	3Cys1Asp/Glu[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	Y
Group 2d	xPRxCGICGxAH	xFDxCSVCTTH	3Cys1Asn[4Fe4S]	3Cys[3Fe4S]	3Cys1His[4Fe4S]	N
Group 3a	xxRxCGxCxxxH	xYDxCxSCATH*	4Cys[4Fe4S]/3Cys1Asp[4Fe4S]	4Cys[4Fe4S]	4Cys[4Fe4S]	Y
Group 3b	xxRICxxCxxxx	xxDPCISCxxH	4Cys[4Fe4S]	4Cys[4Fe4S]	4Cys[4Fe4S]	Y
Group 3c	xxxICGxCxxxH	AYDPCxxCATH*	4Cys[4Fe4S]	4Cys[4Fe4S]	4Cys[4Fe4S]	Y
Group 3d	xxRxCGICPVSH	xxDPCLSCxTH	4Cys[4Fe4S]	Absent	Absent	Y
Group 4a	xxRVCGICGxxH	SLDPCYSCTDR	4Cys[4Fe4S]	Absent	Absent	Y
Group 4b	xERICGICxxxH	SIDPCxSCTxR	4Cys[4Fe4S]	Absent	Absent	Y
Group 4c	xExxCxLCSNxH	SIDPCISCxER	4Cys[4Fe4S]	Absent	Absent	N
Group 4d	xExxCGICSxxH	xxDPCxxCxxR	4Cys[4Fe4S]	Absent	Absent	Y
Group 4e	xxRxCGICSxxH	xIDPCIxCxER	4Cys[4Fe4S]	Absent	Absent	N
Group 4g	xERVCGVCxxSH	SFELCYxCxDR	4Cys[4Fe4S]	Absent	Absent	N
[FeFe]	*P1 motif*	*P2 motif*	*P3 motif*	*Domains*		
Group A1	xTSCCPxWV	xxxPCxxKK	ExMxCxxGCxxGG	M1, M2 or M3		
Group A2	FTSCCPxWx	CxMPCxAKK	ExMACPGGCxxGG	M3		
Group A3	xTSCC/SPxW	xxMPCxAKK	ExMxCxGGCxxGG	M2, M3 or M4		
Group A4	FTSCCPxWV	xxMPCTCKx	EVMxCPxGCxxGG	M2 or M3		
Group B	xTSC[C]CPxxx	FxGPCxAKK	ExMxCxGGCxxGP	M2 or M3		
Group C	IxxxCPxxx	FxxPCxxKx	ExxxCxxGCxxGP	M2P		
[Fe]	*Fe motif*					
HmdI	xTHACTIPT					

Abbreviations: N, no; Y, yes.
The consensus motifs surrounding the metal-ligating cysteine residues are shown for each hydrogenase type: L1 and L2 motifs for [NiFe]-hydrogenases (Vignais and Billoud, 2007); P1, P2 and P3 motifs for [FeFe]-hydrogenases (Vignais and Billoud, 2007); and the Cys176-based binding motif in [Fe]-hydrogenases (Shima *et al*., 2008). For [NiFe]-hydrogenases, the number, configuration and ligands of the iron–sulphur clusters in the small subunit are listed (the consensus metal-binding motifs for these clusters are listed in Supplementary Table S5). 'Extension' refers to whether the C-terminus of the [NiFe]-hydrogenase large subunit is predicted to be cleaved by an endopeptidase during maturation. For [FeFe]-hydrogenases, the domain structures of the catalytic subunit are shown. Building on existing schemes (Vignais and Billoud, 2007; Calusinska *et al*., 2010), we defined the domain organisation of the catalytic subunit as M1 (H-cluster only), M2 (H-cluster, two FeS clusters), M3 (H-cluster, four FeS clusters), M4 (H-cluster, five FeS clusters) or M2P (H-cluster, PAS domain, two FeS clusters) (see also Figure 2 and Supplementary Table S1).

As detailed in Table 2, the content of the hydrogenase metal centres also differed between subgroups, further confirming phylogenetic placements. Cysteine residues that bind the metal ions of the catalytic centres of the three types of hydrogenase are conserved, suggesting the chemical structures of the active site do not vary. However, the neighbouring residues in the motifs binding the [NiFe]-centre (L1, L2) and [FeFe]-centre (P1, P2, P3) vary between subgroups that may influence the catalytic behaviour of the enzymes. The number, configuration and ligands of the iron–sulphur clusters of the [NiFe]-hydrogenase small subunits differ between subgroups (Table 2). Nonstandard ligands for iron–sulphur clusters, that is, Asp, Glu, Asn and His, were common in the clusters proximal and distal to the active site. The number of iron–sulphur clusters associated with the [FeFe]-hydrogenase catalytic domain also varies (Table 2 and Figure 2).

The determinants of hydrogen metabolism are widely distributed in bacterial and archaeal genomes

By curating the sequences of all hydrogenases in public sequence databases, we were able to comprehensively map the distribution and diversity of hydrogenases across sequenced microorganisms. Genes encoding putative hydrogenases were detected in 1397 species (Supplementary Table S3) across 55 phyla (Supplementary Table S4) (note that not all species within these phyla contain hydrogenases). The [NiFe]-hydrogenases were the most widespread of the enzymes, occurring in 36 bacterial and 6 archaeal phyla. Consistent with our current

Group 1: Membrane-bound H_2-uptake [NiFe]-hydrogenases

Group 1a: Ancestral

e.g. *Methanosarcina barkeri* (WP_011366827.1)

Group 1b: Prototypical

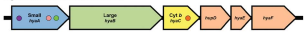
e.g. *Wolinella succinogenes* (WP_011139487.1)

Group 1c: Hyb-type

e.g. *Escherichia coli* (WP_021567771.1)

Group 1d: Oxygen-tolerant

e.g. *Escherichia coli* (WP_021522685.1)



Group 1e: Isp-type

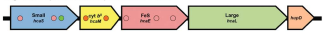
e.g. *Allochromatium vinosum* (WP_012971230.1)

Group 1f: Oxygen-protecting

e.g. *Geobacter sulfurreducens* (WP_010943706.1)

Group 1g: Crenarchaeota-type

e.g. *Thermoproteus neutrophilus* (WP_012340775.1)



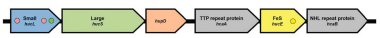
Group 1h/5: Actinobacteria-type

e.g. *Mycobacterium smegmatis* (WP_063984103.1)

Group 2: Cytosolic H_2-uptake [NiFe]-hydrogenases

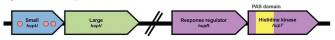
Group 2a: Cyanobacteria-type

e.g. *Mycobacterium smegmatis* (WP_063909634.1)



Group 2b: HK-linked regulatory

e.g. *Ralstonia eutropha* (WP_011153546.1)



Group 2c (putative): DGC-linked regulatory

e.g. *Magnetospirillum magneticum* (WP_011383527.1)

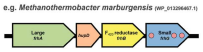
Group 2d: Aquificae-type

e.g. *Aquifex aeolicus* (WP_010880487.1)

Group 3: Cytosolic bidirectional [NiFe]-hydrogenases

Group 3a: F_{420}-coupled

e.g. *Methanothermobacter marburgensis* (WP_013206407.1)

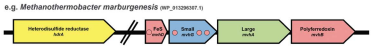


Group 3b: NADP-coupled

e.g. *Pyrococcus furiosus* (WP_011012929.1)

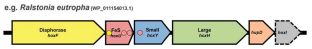
Group 3c: HDR-linked

e.g. *Methanothermobacter marburgensis* (WP_013206387.1)



Group 3d: NAD-coupled

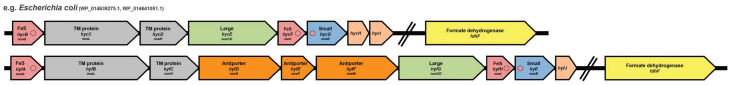
e.g. *Ralstonia eutropha* (WP_011154913.1)



Group 4: Membrane-bound H_2-evolving hydrogenases

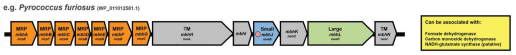
Group 4a: Formate hydrogenlyases

e.g. *Escherichia coli* (NP_014638275.1, WP_014661691.1)



Group 4b: Mrp-linked

e.g. *Pyrococcus furiosus* (WP_011013591.1)



Group 4c: CODH-linked

e.g. *Carboxydothermus hydrogenoformans* (WP_011344721.1)

Group 4d: Polyferredoxin-coupled

e.g. *Methanothermobacter marburgensis* (WP_013206817.1, WP_013206415.1)

Group 4e: Ech-type

e.g. *Methanosarcina barkeri* (WP_011360188.1)

Group 4f (putative): Ehf-type

e.g. *Desulfurosporosinus orientis* (WP_014193792.1)

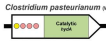
Group A: Prototypical & Bifurcating [FeFe]-hydrogenases

Group A1: Prototypical

e.g. *Chlamydomonas reinhardtii* (XP_001693376.1) (Monomeric, M1)

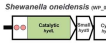
Trichomonas vaginalis (XP_001330775.1) (Monomeric, M2)

Clostridium pasteurianum (WP_014455018.1) (Monomeric, M3)

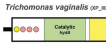


Desulfovibrio vulgaris (WP_010936957.1) (Dimeric, M2)

Shewanella oneidensis (WP_011973665.1) (Trimeric, M2)



Trichomonas vaginalis (XP_001580480.1) (Fusion Protein, M3)

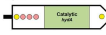


Group A2: Glutamate synthase-linked (putative)

e.g. *Desulfovibrio vulgaris* (WP_010939950.1) (Dimeric, M3)

Group A3: Electron-bifurcating

e.g. *Thermotoga maritima* (WP_004081977.1) (Trimeric, M4)

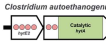


Desulfovibrio fructosivorans (WP_035990063.1) (Tetrameric, M3)

Group A4: Formate dehydrogenase-linked

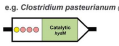
e.g. *Clostridium pasteurianum* (WP_022870234.1) (Trimeric, M2)

Clostridium autoethanogenum (WP_022162897.1) (Hexameric, M3)

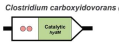


Group B: Ancestral [FeFe]-hydrogenases (putative)

e.g. *Clostridium pasteurianum* (WP_023978023.1) (Monomeric, M2)

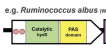


Clostridium carboxydovorans (WP_007661531.1) (Monomeric, M3)

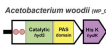


Group C: Sensory [FeFe]-hydrogenases (putative)

e.g. *Ruminococcus albus* (WP_002546614.1) (Phosphatase-linked, M2P)



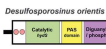
Acetobacterium woodii (WP_014323555.1) (Histidine kinase-linked, M2P)



Geobacter sulfurreducens (WP_014552161.1) (Transcriptional activator-linked, M2P)

Clostridium lundense (WP_027626440.1) (Methyl-accepting chemotaxis protein-linked, M2P)

Desulfosporosinus orientis (WP_014198347.1) (Diguanylate cyclase-linked, M2P)



[Fe]-hydrogenases

e.g. *Methanothermobacter marburgensis* (WP_013206316.1)

knowledge, they were widespread in all classes of *Proteobacteria*, as well as *Firmicutes*, *Cyanobacteria*, *Aquificae*, *Euryarchaeota* and *Crenarchaeota*. [NiFe]-hydrogenases were also common in multiple phyla where hydrogenases have yet to be described, notably *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Planctomycetes* and *Verrucomicrobia*. Putative group A1 [FeFe]-hydrogenases were detected in 12 phyla of anaerobic bacteria, 5 phyla of unicellular eukaryotes and, surprisingly, the single-amplified genome associated to the archaeal candidate phylum *Diapherotrites* (pMC2A384). Other types of [FeFe]-hydrogenases, including bifurcating, ancestral and sensory varieties, were exclusive to anaerobic bacteria such as *Firmicutes*, *Bacteroidetes*, *Spirochaetes*, *Thermotogae* and *Fusobacteria*. Both [NiFe]- and [FeFe]-hydrogenase genes were also identified within most genomes of newly characterised phyla, including the databases of the MDM project (Figure 3). In contrast, genes encoding the functionally restricted [Fe]-hydrogenases were exclusively found within 25 methanogen genomes (Supplementary Figure S6). Phylogenetic distribution correlates with oxygen preference: obligate anaerobes encoded O_2-sensitive [NiFe]- and [FeFe]-hydrogenases; obligate aerobes encoded O_2-tolerant [NiFe]-hydrogenases (groups 1d, 1h, 2a, 3b, 3d); and genomes of facultative anaerobes contained a diverse range (Figure 3). Many hydrogenases have a mosaic distribution that poorly reflects 16S rRNA gene sequence phylogeny, for example, the abundant group 1d and group 3b [NiFe]-hydrogenases (Supplementary Figures S1 and S3), suggesting strong pressure for lateral acquisition of these enzymes.

The determinants of microbial hydrogen metabolism are widely distributed in ecosystems

We subsequently gained insight into the distribution of hydrogenases at the ecosystem level through identifying and analysing hydrogenase sequence reads in 20 publicly available metagenomes (Supplementary Table S5). Sequence reads corresponding to the catalytic subunits/domains of hydrogenases were detected in all metagenome samples analysed, and all 22 [NiFe]-hydrogenase subgroups and 3 [FeFe]-hydrogenase groups (but not the methanogen-specific [Fe]-hydrogenases) were detected in at least two samples each. The normalised abundance of hydrogenase reads ranged approximately ~50-fold, from <0.001% of the total sequence reads in lake and coastal waters to >0.04% reads in permafrost soils and hot springs (Supplementary Table S5). The distribution of hydrogenase-encoding genes in the sequence reads vary depending on the aeration state of the samples: oxic agricultural and forest soils were dominated by aerobically adapted uptake and bidirectional [NiFe]-hydrogenase reads (groups 1d, 1h, 3b, 3d); anoxic termite and human guts contained a high abundance of fermentative and putative sensory [FeFe]-hydrogenase reads (groups A, B, C); and the bog soils, permafrost soils and hot springs contained diverse [NiFe]- and [FeFe]-hydrogenase reads (Figure 4 and Supplementary Figure S7). Hydrogenase-encoding genes were far less abundant within the aquatic ecosystems tested compared with soil and enteric systems; nevertheless, numerous sequence reads homologous to the *Robiginitalea biformata* hydrogenase of the still-uncharacterised group 1g [NiFe]-hydrogenases were detected in deep ocean samples. The quantity and distribution of hydrogenase reads was consistent between paired metagenome samples, including samples from equivalent ecosystems taken at different locations (for example, Atlantic Ocean vs Indian Ocean) (Figure 4). An interesting exception was the termite gut samples, with hydrogenases predicted to be 10 times more abundant in a gut sample of an African termite (*Cubitermes* sp.) compared with an American termite (*Nasutitermes* sp.) (Supplementary Table S5).

Discussion

Molecular hydrogen is a major electron donor for respiration in both anoxic and oxic ecosystems

Molecular hydrogen occurs ubiquitously in the environment, as a result of production from biological, geothermal and atmospheric sources (Schwartz *et al*., 2013). Our analysis suggests microorganisms are capable of respiring this fuel source in a wide variety of ecosystems, ranging from the hypoxic H_2-enriched environments of animal guts and bog soils to aerated soils and waters containing trace concentrations of H_2. Group 1 and 2 [NiFe]-hydrogenases that mediate respiratory H_2 uptake were encoded in 19 of the 20 ecosystems surveyed (Figure 4), and were widely distributed in the bacterial and archaeal phyla (Figure 3). As summarised in Table 1, these enzymes have differentiated into multiple subgroups that differ in their redox couplings, oxygen tolerance, affinities and cellular interactions. This enables these enzymes to support life across a wide range

Figure 2 Genetic organisation of hydrogenases. The genes surrounding the catalytic subunit of representatives of each subtype/subclass are shown to-scale. Genes/domains are colour coded as follows: green = catalytic site; blue = small subunit; yellow = electron acceptor or donor; red = redox subunit; light orange = maturation factor; dark orange = ion-translocation module; purple = regulatory module; grey = conserved hypothetical. Redox-active centres are shown in circles, where: orange = heme; red = [4Fe4S] cluster; yellow = [2Fe2S] cluster; green = [3Fe4S] cluster; purple = [4Fe3S] cluster. Genes are named according to nomenclature if previously defined. There are often variations in the genetic organisation within subgroups, for example, cytochrome *c* subunits replace cytochrome *b* subunits in most group 1a and 1b [NiFe]-hydrogenases in δ-*Proteobacteria*. However, the organisations depicted reflect the most common organisation, as inferred using the Microbial Genomic Context Viewer.

a
Number of Sequenced Representatives
0
50
100
150
200
250
300
350
400
Membrane-Bound H2-Uptake [NiFe] Hydrogenases
Cytosolic H2-Uptake [NiFe] Hydrogenases
Bidirectional [NiFe] Hydrogenases
Membrane-Bound H2-Evolving [NiFe] Hydrogenases
[FeFe]-Hydrogenases
[Fe]-Hydrog-enases
1a: Ancestral
1b: Prototypical
1c: Hyb-type
1d: Oxygen-tolerant
1e: Isp-type
1f: Oxygen-protecting
1g: Crenarchaeota-type
1h/5: Actinobacteria-type
2a: Cyanobacteria-type
2b: HK-linked regulatory
2c: DGC-linked regulatory
2d: Aquificae-type
3a: F420-coupled
3b: NADPH-coupled
3c: HDR-linked
3d: NADH-coupled
4a: Formate hydrogenlyase
4b: Mrp-linked
4c: CODH-linked
4d: Polyferredoxin-coupled
4e: Ech-type
4f: Ehf-type
A1: Prototypical
A2: GS-linked
A3: Bifurcating
A4: FDH-linked
B: Ancestral
C: Sensory
Methylene-H4MPT-coupled
Epsilonproteobacteria
Deltaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Other Bacteria
Cyanobacteria
Bacteroidetes
Firmicutes
Actinobacteria
Eukarya
Other Archaea
Crenarchaeota
Euryarchaeota
b
PHYLUM
No. Species
NiFe Aerobic Uptake
NiFe Anaerobic Uptake
NiFe Bidirec-tional
NiFe Evolving
FeFe Evolving
NiFe Sensory
FeFe Sensory
Fe Uptake
BACTERIA
518 Proteobacteria
309 Firmicutes
149 Actinobacteria
71 Cyanobacteria
59 Bacteroidetes
26 Spirochaetes
18 Thermotogae
15 Chloroflexi
14 Aquificae
14 Synergistetes
12 Fusobacteria
10 Chlorobi
9 Verrucomicrobia
6 Acidobacteria
5 Deferribacteres
5 Planctomycetes
4 Nitrospirae
3 Thermodesulfobacteria
3 Cloacimonetes
2 Tenericutes
2 Atribacteria
2 Chrysiogenetes
2 Deinococcus-Thermus
2 Dictyoglomi
2 Ignavibacteriae
2 Omnitrophica
1 Acetothermia
1 Aerophobetes
1 Aminicenantes
1 Armatimonadetes
1 Calditrix
1 Candidate phylum BRC1
1 Candidate phylum WS1
1 Elusimicrobia
1 Fervidibacteria
1 Fibrobacteres
1 Gemmatimonadetes
1 Hydrogenedentes
1 Latescibacteria
1 Lentisphaerae
1 Marinimicrobia
1 Mollusca
1 Nitrospina
1 Parcubacteria
ARCHAEA
75 Euryarchaeota
23 Crenarchaeota
2 Thaumarchaeota
1 Aenigmarchaeota
1 Algarchaeota
1 Diapherotrites
1 Korarchaeota
EUKARYA
3 Chlorophyta
2 Heterokonta
2 Metamonada
1 Amoebozoa
1 Percolozoa

of ecological niches (Pandelia *et al.*, 2012; Schwartz *et al.*, 2013).

Integrating our analysis with the wider literature, we suggest oxygen partial pressure (pO_2) principally drove the evolution and distribution of respiratory hydrogenases. Phylogenetic analysis reveals the deepest-branching forms of these enzymes (groups 1a, 1b) are O_2 sensitive and mediate anaerobic respiration in strictly anaerobes (Supplementary Figures S1 and S2). Such enzymes are abundant in hypoxic soils (for example, bog soils, permafrost soils) (Supplementary Figure S7), and are

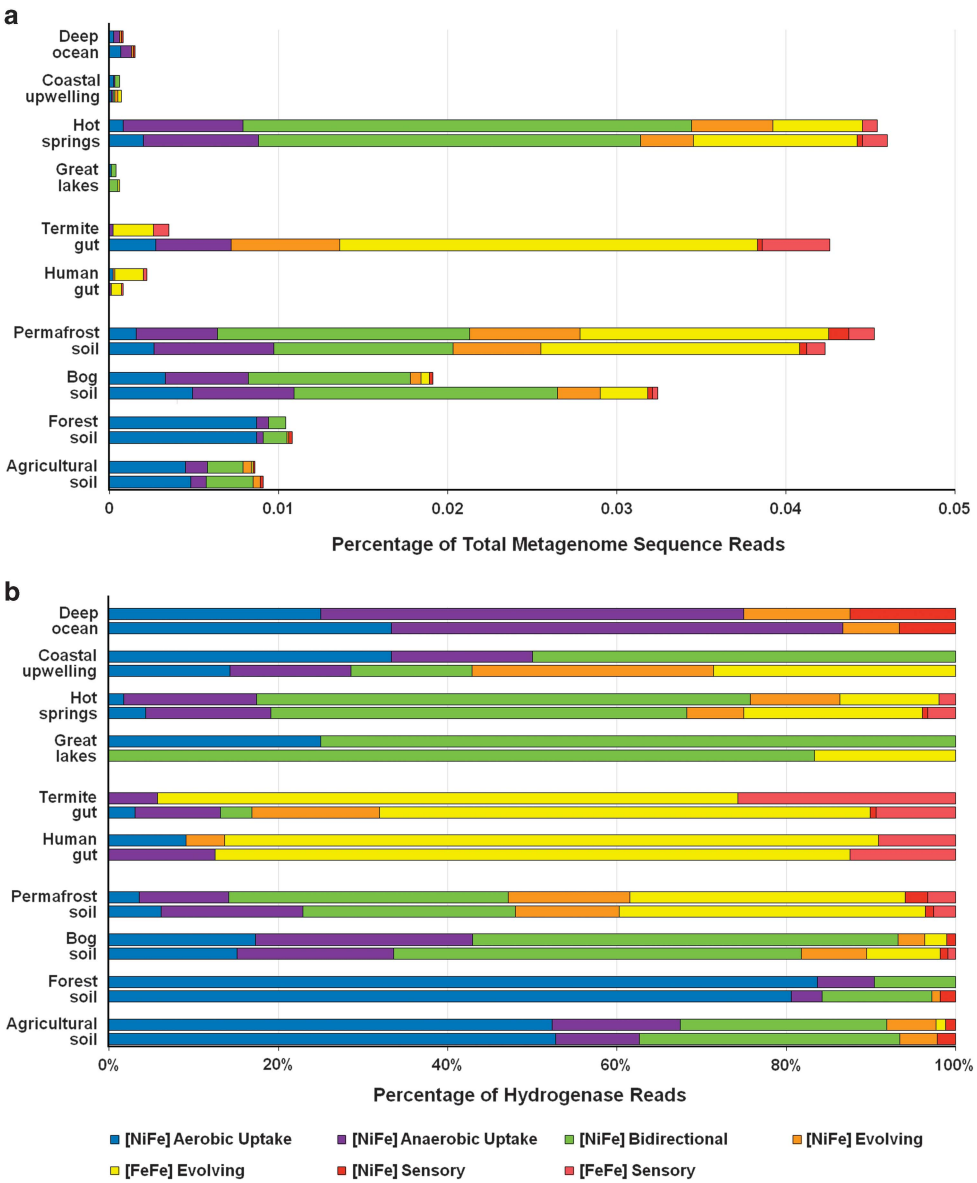


Figure 4 Distribution of hydrogenases in ecosystems. The distribution of different hydrogenase types was analysed in 20 metagenomes. Hydrogenases were subdivided into seven types as described in the legend of Figure 3. Metagenomes were screened using the sequences of the catalytic subunits ([NiFe]-hydrogenases, [Fe]-hydrogenases) or catalytic domains ([FeFe]-hydrogenases) listed in Supplementary Table S1. (**a**) Percentage of sequence reads for each hydrogenase type identified within 1 million random metagenome reads. (**b**) Percentage of sequence reads for each hydrogenase type compared with total hydrogenase sequence reads. Supplementary Figure S5 shows the metagenome distribution by [NiFe]-hydrogenase subgroup and [FeFe]-hydrogenase group. Note that no [Fe]-hydrogenases were detected in these metagenomes.

Figure 3 Distribution of hydrogenases in microorganisms. (**a**) Distribution by hydrogenase type. (**b**) Distribution by phyla. The cells are shaded by the number of hydrogenases detected in each phyla (light = few hydrogenases, dark = many hydrogenases, grey = no hydrogenases). Hydrogenases were subdivided into the following seven types based on their determined or predicted functions: [NiFe] aerobic uptake (groups 1d, 1h, 2a) [NiFe] anaerobic uptake (groups 1a, 1b, 1c, 1e, 1f, 1g, 3a), [NiFe] bidirectional (groups 3b, 3c, 3d), [NiFe] evolving (groups 4a, 4b, 4c, 4d, 4e, 4f), [FeFe] evolving (groups A, B), [NiFe] regulatory (groups 2b, 2c) and [FeFe] regulatory (group C).

predominantly found in the genomes of anaerobic *Firmicutes* and δ-*Proteobacteria* (Supplementary Figure S1) capable of H_2-dependent sulphate reduction, metal reduction and dehalorespiration (Schwartz *et al.*, 2013). The mid-branching heterotetrameric hydrogenases (groups 1c, 1e, 1g) were more taxonomically dispersed (Supplementary Figure S1) and appear to support roles in fumarate and nitrate respiration, anoxygenic photosynthesis and chemolithoautotrophy across a diversity of taxa (Pandelia *et al.*, 2012). In contrast, the more recently branching lineages (groups 1d, 1h and 2a) appear to be oxygen-tolerant enzymes that mediate respiration in aerobes and facultative anaerobes (Table 1). These enzymes were predominant in aerated samples (Figure 4), with 0.004 to 0.009 % of total metagenome sequence reads in agricultural and forest soils corresponding to the group 1h enzyme (Supplementary Table S5). Our analysis suggests that these subgroups independently developed mechanisms to tolerate O_2 following the emergence of oxygenic photosynthesis (Figure 1); consistently, the well-reported proximal 6Cys[4Fe3S] cluster (Fritsch *et al.*, 2011) is exclusive to 1d enzymes (Table 2), indicating the 1h and 2a subgroups use alternative mechanisms (possibly using other non-standard clusters) to prevent or reverse formation of O_2-inhibited states.

Oxygen-tolerant uptake hydrogenases are significantly more widespread than the literature currently reports. Aerobic H_2 uptake has only been reported in three dominant soil phyla to date: many α-, β- and γ-*Proteobacteria* (for example, *Ralstonia eutropha*; Schwartz *et al.*, 2013) can grow chemolithoautotrophically using biologically evolved H_2, whereas certain model *Actinobacteria* (for example, *Mycobacterium smegmatis*; Greening *et al.*, 2014) and *Acidobacteria* (that is, *Pyrinomonas methylaliphatogenes*; Greening *et al.*, 2015a) enhance their persistence by scavenging atmospheric H_2. However, Figure 3 reveals that the group 1d, 1h, and 2a hydrogenases mediating such processes are also encoded in some 17 bacterial and archaeal phyla, among them representatives of all 9 of the most dominant phyla in global soils (Janssen, 2006). Most significantly, the group 1h [NiFe]-hydrogenases that mediate tropospheric H_2 oxidation are encoded in multiple representatives of undercultured, slow-growing phyla (that is, *Acidobacteria*, *Verrucomicrobia*, *Chloroflexi* and *Planctomycetes*). These findings are consistent with our recent hypothesis that H_2 serves as an energy source for the maintenance of dormant soil bacteria (Greening *et al.*, 2015b). Hydrogenase-encoding genes were also identified in the genomes of multiple seemingly obligate methane oxidisers, ammonia oxidisers and nitrite oxidisers (Supplementary Table S1), suggesting H_2 may serve as a fuel source for growth or survival of these bacteria and archaea. In line with this, it was recently demonstrated that *Nitrospira moscoviensis* of the phylum *Nitrospirae* is capable of hydrogenotrophic growth using a group 2a [NiFe]-hydrogenase (Koch *et al.*, 2014). Aerobic H_2 oxidation may therefore provide hitherto-unrecognised metabolic flexibility in microorganisms controlling the methane and nitrogen cycles.

Determinants of fermentative hydrogen production are universally distributed

Since its discovery in the early twentieth century (Stephenson and Stickland, 1932), it has been widely believed that fermentative H_2 evolution occurs exclusively in anaerobic microorganisms. This notion was recently challenged by the discovery that the obligately aerobic soil bacterium *Mycobacterium smegmatis* evolves H_2 using a tightly regulated hydrogenase to maintain redox balance under hypoxia (Berney *et al.*, 2014). The surveys presented in this work demonstrate that homologues of the group 3b [NiFe]-hydrogenase mediating mycobacterial H_2 evolution actually have the most extensive distribution (phylum level) of all the subgroups in our database. Once thought to be confined to anaerobic archaea (Ma *et al.*, 1993), these enzymes actually occur in at least 27 bacterial and archaeal phyla, among them multiple representatives of the 'MDM' (Rinke *et al.*, 2013) (Figure 3). In addition to the versatile group 3d [NiFe]-hydrogenases (Burgdorf *et al.*, 2005), these oxygen-tolerant enzymes are proposed to serve as redox valves that interconvert electrons between NAD(P)H and H_2 depending on the availability of exogenous electron acceptors (Greening and Cook, 2014). These enzymes are also abundant at the metagenome level, constituting dominant groups in aerated soils and hot spring ecosystems (Figure 4). These findings may help to explain why the communities of such ecosystems are relatively stable despite pO_2 fluctuations. Unsurprisingly, the classical determinants of H_2 fermentation were abundant in anoxic ecosystems (Figure 4). Figure 3 shows that formate hydrogenlyases (group 4a [NiFe]-hydrogenases) are widespread in enteric bacteria that adopt a facultatively fermentative lifestyle. The group A1 [FeFe]-hydrogenases, which mediate ferredoxin-dependent H_2 production (Peters *et al.*, 1998), are distributed in numerous obligately fermentative bacteria (for example, clostridia), eukaryotes containing hydrogenosomes (for example, *Trichomonas vaginalis*) and unicellular algae mediating photobiological H_2 production (for example, *Chlamydomonas reinhardtii*). On the basis of domain conservation and phylogenetic similarity, we predict the still-uncharacterised group B [FeFe]-hydrogenases serve a similar function.

Energy-converting and electron-bifurcating complexes enhance the efficiency and flexibility of anaerobe-type hydrogenases

Although group 4 [NiFe]-hydrogenases are traditionally known for their roles in fermentation, the

majority of these enzymes have a respiratory function. They associate into complexes comprising primary dehydrogenases and terminal hydrogenases and conserve the energy liberated during electron transfer as a proton- or sodium-motive force (Buckel and Thauer, 2013). Our analysis shows these enzymes have retained roles in anaerobic microorganisms, especially *Firmicutes*, *Proteobacteria* (γ, δ and ε classes) and methanogens (Figure 3), and contribute to hydrogenase diversity in metagenomes (Figure 4). They appear to have diverse physiological roles, as reflected by their wide-branching phylogeny (Figure 1) and highly modular genetic organisation (Figure 2). This enables them to liberate electrons from low-potential donors, namely formate (group 4a, 4b, possibly 4f), carbon monoxide (group 4b, 4c) or ferredoxin (group 4d, 4e), whereas protons serve as the terminal electron acceptor. Though minimalistic, the respiratory chains they form are often highly efficient and may provide a primary strategy for energy generation within particularly oligotrophic environments; this was emphasised by the recent discovery of a complex in the deep-sea vent archaeon *Thermococcus onnurineus* that sustains growth across a narrow energy bracket by transferring electrons from formate to protons (Kim *et al*., 2010; Lim *et al*., 2014). Others are highly flexible, as demonstrated by the multifaceted roles of the physiologically reversible Ech hydrogenase (group 4e) in hydrogenotrophic vs aceticlastic methanogenesis (Meuer and Kuettner, 2002). Our phylogenetic analyses suggest that the ancestral forms of the group 4 enzymes—and likely [NiFe]-hydrogenases as a whole—may have been formate-oxidising, H_2-evolving, energy-transducing complexes. We discovered a deep-branching lineage of these enzymes in *Firmicutes* (candidate group 4f [NiFe]-hydrogenases) that align closely with the functionally cryptic Ehr complexes (homologues of group 4 [NiFe]-hydrogenases lacking Ni-binding cysteine residues; elaborated on in Marreiros *et al*., 2013) and yet possess the critical cysteine residues required for [NiFe]-centre ligation.

Many of the anaerobe-type hydrogenases we identified are predicted to mediate electron bifurcation, a recently discovered third mode of energy conservation. Electron-bifurcating hydrogenases are bidirectional enzymes that energise the endergonic reaction of the reduction of ferredoxin with H_2 by simultaneously reducing a relatively electropositive acceptor (for example, heterodisulphide, NAD, NADP) (Buckel and Thauer, 2013). The group 3c [NiFe]-hydrogenase in functional complex with heterodisulphide reductase, for example, simultaneously reduces ferredoxin and heterodisulphide during H_2 oxidation (Kaster *et al*., 2011); these enzymes complete the recently elucidated Wolfe cycle of methanogenesis (Thauer, 2012), and are also distributed in some bacteria (for example, δ-*Proteobacteria*) (Figure 4). The group A3 [FeFe]-hydrogenases reversibly bifurcate electrons from H_2 to ferredoxin and NAD using trimeric or tetrameric complexes; in the reverse reaction, energy conserved during the oxidation of ferredoxin is used to drive the thermodynamically unfavourable production of H_2 from NADH (Schut and Adams, 2009; Schuchmann and Müller, 2012). A subtype of the group A4 [FeFe]-hydrogenases can also bifurcate electrons from H_2 to NADP and ferredoxin, and act physiologically in hexameric complexes with formate dehydrogenase (Wang *et al*., 2013). We analysed the genetic organisation of the 705 group A [FeFe]-hydrogenases represented in our database in order to identify putative electron-bifurcating complexes (Figure 2 and Supplementary Table S1). This analysis suggested that whereas putative NADP-dependent bifurcating complexes are rare (7 sequences), putative NAD-dependent bifurcating complexes are very abundant in anaerobic bacteria (391 sequences). The group A3 [FeFe]-hydrogenases are highly flexible, capable of both dissipating excess reductant during fermentation (for example, cellulose fermentation) and generating reduced electrons for carbon fixation (for example, acetogenesis) and respiration (via the sodium-motive ferredoxin-NAD oxidoreductase complex) (Buckel and Thauer, 2013; Schuchmann and Muller, 2014). Supported by PCR amplicon sequencing (Zheng *et al*., 2013), metagenome analysis (Figure 4) demonstrates that group A [FeFe]-hydrogenases, including probable bifurcating hydrogenases, are abundant in termite guts.

Hydrogen sensing may be more important than previously recognised

Our exploration of hydrogenase sequences across a diversity of environments also uncovered evidence that hydrogen-based signal transduction cascades are more significant than previously anticipated. The only characterised sensory hydrogenases to date are the group 2b [NiFe]-hydrogenases (for example, *Ralstonia eutropha*, *Rhodobacter capsulatus*) (Lenz and Friedrich, 1998; Vignais *et al*., 2005); these enzymes have adapted the [NiFe] active site to sense high partial pressures of H_2 and in turn activate two-component regulatory cascades that control expression of respiratory hydrogenases (Greening and Cook, 2014). Our analysis shows these enzymes are restricted to *Proteobacteria* (α, β and γ classes) (Figure 3 and Supplementary Figure S2) and are present in soil environments characterised by logarithmic variations in pH_2 (Figure 4 and Supplementary Figure S7). We identified a sister lineage, the group 2c [NiFe]-hydrogenases, that appear to be co-transcribed with diguanylate cyclases/phosphodiesterases (Figure 2). Through modulation of cyclic di-GMP production, we hypothesise these enzymes regulate global cellular functions during adaptation to H_2-rich vs H_2-deprived environments; currently, however, H_2-dependent signal cascades have only been shown to regulate the expression of other

hydrogenases. Our genome and metagenome surveys suggest these enzymes are rare (Figures 3 and 4), and are primarily found in methane-oxidising bacteria and sulphate-reducing bacteria that inhabit aquatic environments (Supplementary Figure S2). The presence of helix-turn-helix protein-encoding genes immediately downstream of group 2 [NiFe]-hydrogenase genes in some *Aquificae* and *Crenarchaeota* (Supplementary Figure S2) is also suggestive of a regulatory role and requires further study.

Looking more widely, it is probable that the group C [FeFe]-hydrogenases of anaerobic bacteria have a sensory role. Our conserved domain analysis suggests these enzymes are expressed or fused with putative regulatory components, namely serine/threonine phosphatases, histidine kinases, AAA+-type transcriptional activators, methyl-accepting chemotaxis proteins and again diguanylate cyclases/phosphdiesterases (Figure 2). As with the group 2b and 2c [NiFe]-hydrogenases, the operons encoding these hydrogenases contain predicted PAS domains (Figure 2) that likely transduce the signal of hydrogenase activity to downstream components via a redox-active heme. There is some transcriptional evidence that the putative phosphatase-linked sensory hydrogenases of *Thermoanaerobacterium saccharolyticum* (Shaw *et al*., 2009) and *Ruminococcus albus* (Zheng *et al*., 2014) regulate the transcription of group A [FeFe]-hydrogenases, but it has yet to be biochemically confirmed that these enzymes have a regulatory role. Other sensory hydrogenases may regulate wider cellular functions (for example, motility) in response to changes of pH_2 in anoxic environments. Group C [FeFe]-hydrogenases are abundant in strictly anaerobic bacteria of the phyla *Firmicutes*, *Bacteroidetes*, *Spirochaetes* and *Thermotogae*. These hydrogenases are highly abundant in termite guts and strongly associated with group A [FeFe]-hydrogenases (Supplementary Figure S7).

Conclusions

The surveys reported here suggest that hydrogenases are highly diverse, ancient and widespread. Our work collectively supports the hypothesis that H_2 serves as a widely utilised energy source for microbial growth and survival. Before this study, it was already well established that H_2 metabolism played major roles in certain specific microorganisms and ecosystems (Schwartz *et al*., 2013). However, by comprehensively surveying the distribution of hydrogenases, we have provided evidence that microbial H_2 metabolism is significantly more extensive and elaborate than previously anticipated. Integrating analysis of primary phylogeny, genetic organisation and metal-binding motifs, we demonstrate that hydrogenases have evolved into numerous functionally distinct subgroups/subtypes. This diversification has enabled the primordial process of H_2 metabolism to sustain roles in most major phyla and ecosystems.

We showed that some 51 bacterial and archaeal phyla have the genetic capacity to oxidise or evolve H_2—vastly more than the 13 phyla experimentally shown to metabolise H_2 (Schwartz *et al*., 2013; Greening *et al*., 2015a)—and emphasised through metagenome analysis that microbial H_2 metabolism is likely to be highly important in both oxic and anoxic environments. Our analysis has emphasised that the evolution and distribution of hydrogenases is particularly influenced by pO_2; however, other factors such as pH_2, pH, temperature and metal ion availability are also likely to be profoundly significant (Schwartz *et al*., 2013; Greening and Cook, 2014). However, experimental studies are required to gain a deeper understanding of the ecological significance of H_2 oxidation and evolution.

In light of this work, there are now numerous new avenues to investigate microbial hydrogen metabolism at the microscopic and macroscopic levels: What are the functions of multiple newly defined types of [NiFe]-hydrogenase (groups 1e, 1g, 2c, 2d, 4f) and [FeFe]-hydrogenases (groups A2, B, C)? Why are hydrogenases found in the genomes of microorganisms as diverse as *Acidobacteria*, *Chlorobi*, *Crenarchaeota* and *Bacteroidetes*? What environmental and physiological signals lead to the regulation of the genetic determinants of hydrogen metabolism? How does microbial H_2 metabolism influence anthropogenic ecosystems (for example, wastewater treatment) and how can the reported diversity of hydrogenases be exploited for bioremediation, biofuel production and fuel cell development? How does microbial H_2 metabolism influence community structuring and biogeochemical cycling in soil and aquatic environments? Some of these research questions will be addressed as we further investigate the physiological roles of the hydrogenases described here and the influence of H_2 metabolism in different ecosystems.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank David Duncan, Jeffrey Blanchard, William Mohn, Janet Jansson, Andreas Brune, Vincent Denef, Donald Bryant, Christopher Francis and Silvia G Acinas for sequencing the metagenomes analysed here. We additionally thank the anonymous reviewers for their helpful suggestions. This work was supported by a CSIRO Office of the Chief Executive Postdoctoral Fellowship awarded to CG. MBS and CRC were supported by the Core-funded programmes: Geothermal Resources of New Zealand (GNS Science) and Environmental Technologies (Scion).

References

Afting C, Hochheimer A, Thauer RK. (1998). Function of H_2-forming methylenetetrahydromethanopterin dehydrogenase from *Methanobacterium thermoautotrophicum* in coenzyme F_{420} reduction with H_2. *Arch Microbiol* **169**: 206–210.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Andrews SC, Berks BC, McClay J, Ambler A, Quail MA, Golby P *et al.* (1997). A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology* **143**: 3633–3647.

Barz M, Beimgraben C, Staller T, Germer F, Opitz F, Marquardt C *et al.* (2010). Distribution analysis of hydrogenases in surface waters of marine and freshwater environments. *PLoS One* **5**: e13846.

Beimgraben C, Gutekunst K, Opitz F, Appel J. (2014). hypD as a marker for [NiFe]-hydrogenases in microbial communities of surface waters. *Appl Environ Microbiol* **80**: 3776–3782.

Berney M, Greening C, Conrad R, Jacobs WR, Cook GM. (2014). An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc Natl Acad Sci USA* **111**: 11479–11484.

Brugna-Guiral M, Tron P, Nitschke W, Stetter K-O, Burlat B, Guigliarelli B *et al.* (2003). [NiFe] hydrogenases from the hyperthermophilic bacterium *Aquifex aeolicus*: properties, function, and phylogenetics. *Extremophiles* **7**: 145–157.

Buckel W, Thauer RK. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**: 94–113.

Burgdorf T, van der Linden E, Bernhard M, Yin QY, Back JW, Hartog AF *et al.* (2005). The soluble NAD^+-reducing [NiFe]-hydrogenase from *Ralstonia eutropha* H16 consists of six subunits and can be specifically activated by NADPH. *J Bacteriol* **187**: 3122–3132.

Calusinska M, Happe T, Joris B, Wilmotte A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* **156**: 1575–1588.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. (2008). BLAST+: architecture and applications. *BMC Bioinform* **10**: 421.

Constant P, Chowdhury SP, Hesse L, Pratscher J, Conrad R. (2011). Genome data mining and soil survey for the novel Group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity H2-oxidizing bacteria. *Appl Environ Microbiol* **77**: 6027–6035.

Constant P, Chowdhury SP, Pratscher J, Conrad R. (2010). Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environ Microbiol* **12**: 821–829.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. (2004). WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Deppenmeier U, Blaut M. (1995). Analysis of the vhoGAC and vhtGAC operons from *Methanosarcina mazei* strain Gö1, both encoding a membrane-bound hydrogenase and a cytochrome b. *Eur J Biochem* **269**: 261–269.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Fox JD, Kerby RL, Roberts GP, Ludden PW. (1996). Characterization of the CO-induced, CO-tolerant hydrogenase from Rhodospirillum rubrum and the gene encoding the large subunit of the enzyme. *J Bacteriol* **178**: 1515–1524.

Fritsch J, Scheerer P, Frielingsdorf S, Kroschinsky S, Friedrich B, Lenz O *et al.* (2011). The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature* **479**: 249–252.

Garcin E, Vernede X, Hatchikian EC, Volbeda A, Frey M, Fontecilla-Camps JC. (1999). The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center. *Structure* **7**: 557–566.

Greening C, Berney M, Hards K, Cook GM, Conrad R. (2014). A soil actinobacterium scavenges atmospheric H2 using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proc Natl Acad Sci USA* **111**: 4257–4261.

Greening C, Carere CR, Rushton-Green R, Harold LK, Hards K, Taylor MC *et al.* (2015a). Persistence of the dominance soil phylum *Acidobacteria* by trace gas scavenging. *Proc Natl Acad Sci USA* **112**: 10497–10502.

Greening C, Constant P, Hards K, Morales SE, Oakeshott JG, Russell RJ *et al.* (2015b). Atmospheric hydrogen scavenging: from enzymes to ecosystems. *Appl Environ Microbiol* **81**: 1190–1199.

Greening C, Cook GM. (2014). Integration of hydrogenase expression and hydrogen sensing in bacterial cell physiology. *Curr Opin Microbiol* **18**: 30–38.

Gross R, Simon J, Theis F, Kröger A. (1998). Two membrane anchors of *Wolinella succinogenes* hydrogenase and their function in fumarate and polysulfide respiration. *Arch Microbiol* **170**: 50–58.

Happe T, Naber JD. (1993). Isolation, characterization and N-terminal amino acid sequence of hydrogenase from the green alga Chlamydomonas reinhardtii. *Eur J Biochem* **214**: 475–481.

Horch M, Lauterbach L, Mroginski MA, Hildebrandt P, Lenz O, Zebger I. (2015). Reversible active site sulfoxygenation can explain the oxygen tolerance of a NAD^+-reducing [NiFe] hydrogenase and its unusual infrared spectroscopic properties. *J Am Chem Soc* **137**: 2555–2564.

Houchins JP, Burris RH. (1981). Occurrence and localization of two distinct hydrogenases in the heterocystous cyanobacterium *Anabaena* sp. strain 7120. *J Bacteriol* **146**: 209–214.

Janssen PH. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* **72**: 1719–1728.

Kanai T, Matsuoka R, Beppu H, Nakajima A, Okada Y, Atomi H *et al.* (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J Bacteriol* **193**: 3109–3116.

Kaserer H. (1906). Die Oxydation des Wasserstoffes durch Mikroorganismen. *Centr Bakteriol Parasitenk* **16**: 681–696.

Kaster A-K, Moll J, Parey K, Thauer RK. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc Natl Acad Sci USA* **108**: 2981–2986.

Kim YJ, Lee HS, Kim ES, Bae SS, Lim JK, Matsumi R *et al.* (2010). Formate-driven growth coupled with H_2 production. *Nature* **467**: 352–355.

Koch H, Galushko A, Albertsen M, Schintlmeister A, Gruber-Dorninger C, Lucker S *et al.* (2014). Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**: 1052–1054.

Kulkarni G, Kridelbaugh DM, Guss AM, Metcalf WW. (2009). Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*. *Proc Natl Acad Sci USA* **106**: 15915–15920.

Lane N, Allen JF, Martin W. (2010). How did LUCA make a living? Chemiosmosis in the origin of life. *BioEssays* **32**: 271–280.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinforma* **23**: 2947–2948.

Laska S, Lottspeich F, Kletzin A. (2003). Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiol* **149**: 2357–2371.

Lenz O, Friedrich B. (1998). A novel multicomponent regulatory system mediates H2 sensing in *Alcaligenes eutrophus*. *Proc Natl Acad Sci USA* **95**: 12474–12479.

Leul M, Normand P, Sellstedt A. (2007). The organization, regulation and phylogeny of uptake hydrogenase genes in Frankia. *Physiol Plant* **130**: 464–470.

Lie TJ, Costa KC, Lupa B, Korpole S, Whitman WB, Leigh JA. (2012). Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. *Proc Natl Acad Sci USA* **109**: 15473–15478.

Lim JK, Mayer F, Kang SG, Müller V. (2014). Energy conservation by oxidation of formate to carbon dioxide and hydrogen via a sodium ion current in a hyperthermophilic archaeon. *Proc Natl Acad Sci USA* **111**: 11497–11502.

Lubitz W, Ogata H, Rüdiger O, Reijerse E. (2014). Hydrogenases. *Chem Rev* **114**: 4081–4148.

Lukey MJ, Parkin A, Roessler MM, Murphy BJ, Harmer J, Palmer T *et al.* (2010). How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J Biol Chem* **285**: 3928–3938.

Ma K, Schicho R, Kelly R, Adams MW. (1993). Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci USA* **90**: 5341–5344.

Maier L, Vyas R, Cordova CD, Lindsay H, Sebastian T, Schmidt B *et al.* (2013). Microbiota-derived hydrogen fuels *Salmonella Typhimurium* invasion of the gut ecosystem. *Cell Host Microbe* **14**: 641–651.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C *et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**: D225–D229.

Marreiros BC, Batista AP, Duarte AMS, Pereira MM. (2013). A missing link between complex I and group 4 membrane-bound [NiFe] hydrogenases. *Biochim Biophys Acta* **1827**: 198–209.

Marques MC, Coelho R, De Lacey AL, Pereira AC, Matias PM. (2010). The three-dimensional structure of [NiFeSe] hydrogenase from *Desulfovibrio vulgaris* Hildenborough: a hydrogenase without a bridging ligand in the active site in its oxidised, 'as-isolated' state. *J Mol Biol* **396**: 893–907.

McDowall JS, Murphy BJ, Haumann M, Palmer T, Armstrong FA, Sargent F. (2014). Bacterial formate hydrogenlyase complex. *Proc Natl Acad Sci USA* **111**: E3948–E3956.

Meuer J, Kuettner H. (2002). Genetic analysis of the archaeon Methanosarcina barkeri Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc Natl Acad Sci USA*.

Mills DJ, Vitt S, Strauss M, Shima S, Vonck J. (2013). De novo modeling of the F_{420}-reducing [NiFe]-hydrogenase from a methanogenic archaeon by cryo-electron microscopy. *Elife* **2**: e00218.

Morais-Silva FO, Santos CI, Rodrigues R, Pereira IAC, Rodrigues-Pousada C. (2013). Roles of HynAB and Ech, the only two hydrogenases found in the model sulfate reducer *Desulfovibrio gigas*. *J Bacteriol* **195**: 4753–4760.

Morales SE, Holben WE. (2011). Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol Ecol* **75**: 2–16.

Nicolet Y, Piras C, Legrand P, Hatchikian CE, Fontecilla-Camps JC. (1999). *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* **7**: 13–23.

Ogata H, Kellers P, Lubitz W. (2010). The crystal structure of the [NiFe] hydrogenase from the photosynthetic bacterium Allochromatium vinosum: characterization of the oxidized enzyme (Ni-A state). *J Mol Biol* **402**: 428–444.

Olson JW, Maier RJ. (2002). Molecular hydrogen as an energy source for *Helicobacter pylori*. *Science* **298**: 1788–1790.

Overmars L, Kerkhoven R, Siezen RJ, Francke C. (2013). MGcV: the microbial genomic context viewer for comparative genome analysis. *BMC Genomics* **14**: 209.

Pandelia ME, Lubitz W, Nitschke W. (2012). Evolution and diversification of Group 1 [NiFe] hydrogenases. Is there a phylogenetic marker for O_2-tolerance? *Biochim Biophys Acta - Bioenerg* **1817**: 1565–1575.

Peters JW, Lanzilotta WN, Lemon BJ, Seefeldt LC. (1998). X-ray crystal structure of the Fe-only hydrogenase (CpI) from Clostridium pasteurianum to 1.8 Angstrom resolution. *Science* **282**: 1853–1858.

Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB *et al.* (2014). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta—Mol Cell Res*; doi:10.1016/j.bbamcr.2014.11.021.

Pihl T, Black L, Schulman B, Maier R. (1989). Characterization of hydrogen-uptake activity in the hyperthermophile *Pyrodictium brockii*. *Proc Natl Acad Sci USA* **86**: 138–141.

Porat I, Kim W, Hendrickson EL, Xia Q, Zhang Y, Wang T *et al.* (2006). Disruption of the operon Encoding Ehb hydrogenase limits anabolic CO_2 assimilation in the archaeon *Methanococcus maripaludis*. *J Bacteriol* **188**: 1373–1380.

Pruitt KD, Tatusova T, Maglott DR. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61–D65.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Rákhely G, Kovács ÁT, Maróti G, Barna D, Csanádi G, Latinovics D *et al.* (2004). Cyanobacterial-type, heteropentameric, NAD+-reducing NiFe hydrogenase in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Appl Environ Microbiol* **70**: 722–728.

Saitou N, Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Sapra R, Bagramyan K, Adams MWW. (2003). A simple energy-conserving system: Proton reduction coupled to proton translocation. *Proc Natl Acad Sci USA* **100**: 7545–7550.

Sawers RG, Jamieson DJ, Higgins CF, Boxer DH. (1986). Characterization and physiological roles of membrane-bound hydrogenase isoenzymes from *Salmonella typhimurium*. *J Bacteriol* **168**: 398–404.

Schuchmann K, Muller V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**: 809–821.

Schuchmann K, Müller V. (2012). A bacterial electron-bifurcating hydrogenase. *J Biol Chem* **287**: 31165–31171.

Schut GJ, Adams MW. (2009). The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**: 4451–4457.

Schwartz E, Fritsch J, Friedrich B. (2013). *H2-Metabolizing Prokaryotes*. Springer Berlin Heidelberg: Berlin, Heidelberg.

Schäfer C, Friedrich B, Lenz O. (2013). Novel, oxygen-insensitive group 5 [NiFe]-hydrogenase in Ralstonia eutropha. *Appl Environ Microbiol* **79**: 5137–5145.

Shaw AJ, Hogsett DA, Lynd LR. (2009). Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. *J Bacteriol* **191**: 6457–6464.

Shima S, Pilak O, Vogt S, Schick M, Stagni MS, Meyer-Klaucke W *et al.* (2008). The crystal structure of [Fe]-Hydrogenase reveals the geometry of the active site. *Science* **321**: 572–575.

Shomura Y, Yoon K-S, Nishihara H, Higuchi Y. (2011). Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. *Nature* **479**: 253–256.

Soboh B, Linder D, Hedderich R. (2002). Purification and catalytic properties of a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus hydrogenoformans*. *Eur J Biochem* **269**: 5712–5721.

Soboh B, Linder D, Hedderich R. (2004). A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. *Microbiology* **150**: 2451–2463.

Stephenson M, Stickland LH. (1931). Hydrogenase: a bacterial enzyme activating molecular hydrogen: the properties of the enzyme. *Biochem J* **25**: 205–214.

Stephenson M, Stickland LH. (1932). Hydrogenlyases: bacterial enzymes liberating molecular hydrogen. *Biochem J* **26**: 712–724.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tengölics R, Mészáros L, Győri E, Doffkay Z, Kovács KL, Rákhely G. (2014). Connection between the membrane electron transport system and Hyn hydrogenase in the purple sulfur bacterium, *Thiocapsa roseopersicina* BBS. *Biochim Biophys Acta* **1837**: 1691–1698.

Thauer RK. (2012). The Wolfe cycle comes full circle. *Proc Natl Acad Sci US A* **109**: 15084–15085.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554–557.

Tremblay P-L, Lovley DR. (2012). Role of the NiFe hydrogenase Hya in oxidative stress defense in *Geobacter sulfurreducens*. *J Bacteriol* **194**: 2248–2253.

Vignais P, Elsen S, Colbeau A. (2005). Transcriptional regulation of the uptake (NiFe) hydrogenase genes in Rhodobacter capsulatus. *Biochem Soc Trans* **33**: 28–32.

Vignais PM, Billoud B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**: 4206–4272.

Vignais PM, Billoud B, Meyer J. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* **25**: 455–501.

Volbeda A, Charon MH, Piras C, Hatchikian EC, Frey M, Fontecilla-Camps JC. (1995). Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**: 580–587.

Volbeda A, Darnault C, Parkin A, Sargent F, Armstrong FA, Fontecilla-Camps JC. (2013). Crystal structure of the O2-tolerant membrane-bound hydrogenase 1 from *Escherichia coli* in complex with its cognate cytochrome b. *Structure* **21**: 184–190.

Wang S, Huang H, Kahnt J, Mueller AP, Köpke M, Thauer RK. (2013). NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in Clostridium autoethanogenum grown on CO. *J Bacteriol* **195**: 4373–4386.

Wootton JC, Federhen S. (1996). Analysis of compositionally biased regions in sequence databases. *Methods Enzymol* **266**: 554–571.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al.* (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**: 1661–1665.

Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN *et al.* (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**: 1056–1060.

Wu LF, Mandrand MA. (1993). Microbial hydrogenases: primary structure, classification, signatures and phylogeny. *FEMS Microbiol Rev* **10**: 243–269.

Zheng H, Bodington D, Zhang C, Miyanaga K, Tanji Y, Hongoh Y *et al.* (2013). Comprehensive phylogenetic diversity of [FeFe]-hydrogenase genes in termite gut microbiota. *Microbes Environ* **28**: 491–494.

Zheng Y, Kahnt J, Kwon IH, Mackie RI, Thauer RK. (2014). Hydrogen formation and its regulation in *Ruminococcus albus*: involvement of an electron-bifurcating [FeFe]-hydrogenase, of a non-electron-bifurcating [FeFe]-hydrogenase, and of a putative hydrogen-sensing [FeFe]-hydrogenase. *J Bacteriol* **196**: 3840–3852.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)