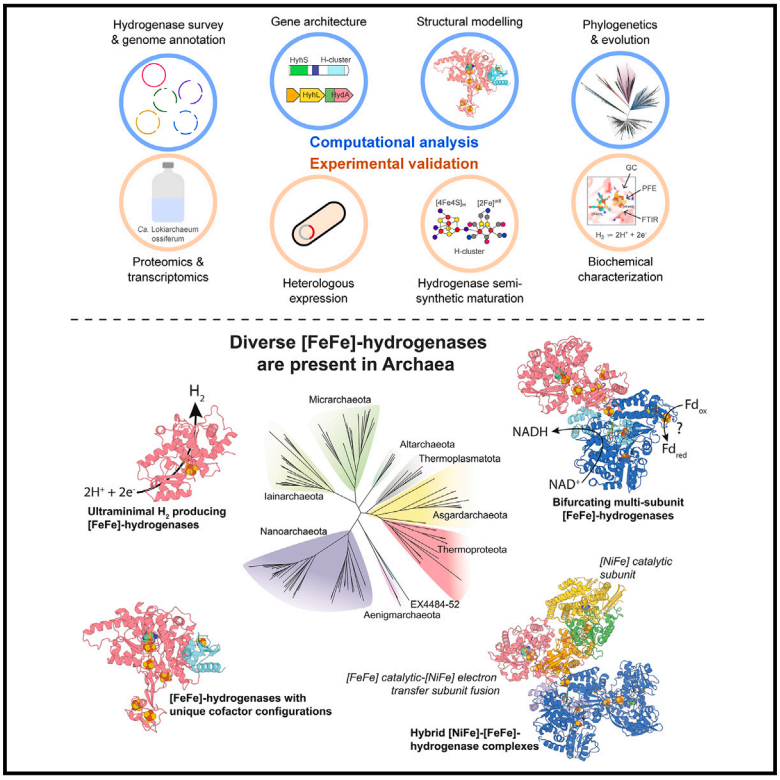
Cell

Minimal and hybrid hydrogenases are active from archaea

Graphical abstract



Highlights

- Archaea from nine different phyla encode structurally diverse [FeFe] hydrogenases
- Active ultraminimal [FeFe] hydrogenases are produced by uncultured DPANN archaea
- Ancient hybrid [FeFe] and [NiFe] hydrogenases are encoded by diverse archaea
- Hydrogen-producing Asgard archaeal cultures express [FeFe] hydrogenases

Authors

Chris Greening, Princess R. Cabotaje, Luis E. Valentin Alvarado, ..., Anja Spang, Jillian F. Banfield, Gustav Berggren

Correspondence

chris.greening@monash.edu (C.G.), rhys.grinter@monash.edu (R.G.), anja.spang@nioz.nl (A.S.), jbanfield@berkeley.edu (J.F.B.), gustav.berggren@kemi.uu.se (G.B.)

In brief

Archaea produce [FeFe] hydrogenases, previously thought to be only made by bacteria and eukaryotes. These enzymes are remarkably diverse, spanning ultraminimal fermentative enzymes to ancient hybrid complexes, and have significant biotechnological potential.

Greening et al., 2024, Cell *187*, 3357–3372
June 20, 2024 © 2024 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2024.05.032

CellPress

Cell

CellPress
OPEN ACCESS

Article

Minimal and hybrid hydrogenases are active from archaea

Chris Greening,[1,2,10,11,*] Princess R. Cabotaje,[3,10] Luis E. Valentin Alvarado,[4,10] Pok Man Leung,[1,2,10] Henrik Land,[3] Thiago Rodrigues-Oliveira,[5] Rafael I. Ponce-Toledo,[5] Moritz Senger,[3] Max A. Klamke,[3] Michael Milton,[1] Rachael Lappan,[1,2] Susan Mullen,[4] Jacob West-Roberts,[4] Jie Mao,[1,6] Jiangning Song,[6] Marie Schoelmerich,[4] Courtney W. Stairs,[7] Christa Schleper,[5] Rhys Grinter,[1,*] Anja Spang,[8,9,*] Jillian F. Banfield,[1,4,*] and Gustav Berggren[3,*]

[1]Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia
[2]SAEF: Securing Antarctica's Environmental Future, Monash University, Clayton, VIC, Australia
[3]Department of Chemistry – Ångström Laboratory, Uppsala University, Uppsala, Sweden
[4]Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94709, USA
[5]Department of Functional and Evolutionary Ecology, Archaea Biology and Ecogenomics Unit, University of Vienna, Vienna, Austria
[6]Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia
[7]Department of Biology, Lund University, Lund, Sweden
[8]Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Den Hoorn, the Netherlands
[9]Department of Evolutionary and Population Biology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, the Netherlands
[10]These authors contributed equally
[11]Lead contact
*Correspondence: chris.greening@monash.edu (C.G.), rhys.grinter@monash.edu (R.G.), anja.spang@nioz.nl (A.S.), jbanfield@berkeley.edu (J.F.B.), gustav.berggren@kemi.uu.se (G.B.)
https://doi.org/10.1016/j.cell.2024.05.032

SUMMARY

Microbial hydrogen (H_2) cycling underpins the diversity and functionality of diverse anoxic ecosystems. Among the three evolutionarily distinct hydrogenase superfamilies responsible, [FeFe] hydrogenases were thought to be restricted to bacteria and eukaryotes. Here, we show that anaerobic archaea encode diverse, active, and ancient lineages of [FeFe] hydrogenases through combining analysis of existing and new genomes with extensive biochemical experiments. [FeFe] hydrogenases are encoded by genomes of nine archaeal phyla and expressed by H_2-producing Asgard archaeon cultures. We report an ultraminimal hydrogenase in DPANN archaea that binds the catalytic H-cluster and produces H_2. Moreover, we identify and characterize remarkable hybrid complexes formed through the fusion of [FeFe] and [NiFe] hydrogenases in ten other archaeal orders. Phylogenetic analysis and structural modeling suggest a deep evolutionary history of hybrid hydrogenases. These findings reveal new metabolic adaptations of archaea, streamlined H_2 catalysts for biotechnological development, and a surprisingly intertwined evolutionary history between the two major H_2-metabolizing enzymes.

INTRODUCTION

Molecular hydrogen (H_2) is heralded as a future green energy carrier. In a biological context, this energy-rich gas already plays a central role in bioenergetics, and evolution has driven an elaborate H_2 economy. Numerous bacteria, archaea, and microbial eukaryotes consume and produce H_2 gas ($H_2 \rightleftharpoons 2\,H^+ + 2\,e^-$) using metalloenzymes called hydrogenases.[1–3] Three hydrogenases have independently evolved in microorganisms, namely the [FeFe], [NiFe], and [Fe] hydrogenases, which differ in their metal cofactors and catalytic mechanism.[4–7] H_2 serves multiple roles in microbial physiology. Microorganisms spanning all three domains produce H_2 to dispose of electrons during fermentation. Numerous bacteria and archaea also use electrons derived from H_2 oxidation for respiration and carbon fixation.[1,2,8–10] More recently, electron-bifurcating hydrogenase complexes have been discovered that are critical for energy conservation in obligate anaerobes.[11–17] Hydrogenases are now recognized as environmentally ubiquitous and taxonomically widespread, encoded in genomes from most bacterial and archaeal phyla as well as many unicellular eukaryotes.[1,3,18] It is increasingly recognized that microbial H_2 metabolism shapes global biogeochemical cycling,[19,20] supports biodiversity of diverse ecosystems,[21,22] and influences health and disease.[23,24] In addition, these efficient enzymes have growing industrial applications in the developing H_2 economy[25,26] and serve as an inspiration for the design of synthetic catalysts.[27,28] H_2 was likely the primordial electron donor but continues to have a central role in microbiology,

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

both as a desirable energy source and as a diffusible electron sink.[29,30] Moreover, it is proposed that H_2 exchange between bacteria and archaea underlies eukaryogenesis, as described in various syntrophy hypotheses.[31–36]

The three hydrogenase classes differ in their physiological roles and taxonomic distribution. [FeFe] hydrogenases are typically fast acting but oxygen sensitive, and are best known for their roles in obligate anaerobes. These enzymes currently comprise four phylogenetically distinct groups (groups A–D), which can be further subdivided through two different schemes based on domain architecture and genetic organization.[37–39] They include monomeric enzymes that couple ferredoxin oxidation to fermentative H_2 production (group A1),[5] trimeric enzymes that reversibly bifurcate electrons from H_2 to nicotinamide adenine dinucleotide (NAD^+) and ferredoxin (group A3),[11] filamentous complexes with formate dehydrogenase that catalyze H_2-dependent CO_2 conversion to formate (group A4),[40] putative sensory hydrogenases in which the catalytic hydrogenase domain is fused with a PAS domain (group C),[41] and several functionally undefined groups (e.g., groups B and D).[1,42] Despite this diversity, [FeFe] hydrogenases are all predicted to rely on the same organometallic cofactor for catalysis, the "H-cluster." To date, these enzymes have been exclusively characterized in anaerobic bacteria and eukaryotes and appear to be absent in cultured archaea.[1,3,37,43] [NiFe] hydrogenases are extraordinarily structurally and functionally diverse enzymes, encoded by bacteria and archaea across all ecosystems. They are presently subdivided into four major groups (groups 1–4) and 29 subgroups that each differ in their phylogeny, genetic organization, and physiological roles.[1,8,44] The catalytic (large) subunit and electron-relaying iron-sulfur (small) subunit of the [NiFe] hydrogenase associate with other subunits, depending on the subgroup; the different complexes formed can mediate respiration, fermentation, energy-conversion, electron bifurcation, carbon fixation, and H_2 sensing processes.[1] In contrast, [Fe] hydrogenases are a much narrower lineage that contribute to archaeal methanogenesis.[6,45] The three hydrogenase classes are phylogenetically unrelated, despite having some similar structural features, and are not thought to genetically or structurally associate. [NiFe] hydrogenases are predicted to have been present in the last universal common ancestor (LUCA), whereas [FeFe] hydrogenases are proposed to have evolved later in fermentative bacteria.[1,30]

In the 8 years since hydrogenase distribution has been comprehensively surveyed,[1] there has been a massive expansion of known microbial diversity—primarily due to genome-resolved metagenomics (i.e., the recovery of microbial genomes from mixed community samples).[46–48] This expansion has been particularly pronounced in the domain Archaea.[49] Notable developments include the discovery of the Asgard superphylum from which eukaryotes are predicted to have evolved,[50–54] the rapid expansion of the DPANN superphylum that includes obligate symbionts,[47,55–57] and the discovery of new lineages capable of anaerobic alkane metabolism.[58–61] Most of these archaea are uncultivated, necessitating culture-independent approaches to characterize them. Many novel lineages appear to be capable of H_2 metabolism. For example, the recently cultured Asgard archaeon "*Candidatus* Prometheoarchaeum syntrophicum" fermentatively generates H_2, seemingly through the activity of an electron-bifurcating/confurcating [NiFe] hydrogenase, which is consumed by its syntrophic methanogenic partner.[35] The H_2-metabolizing enzymes reported in these newly discovered archaea fall into a range of established and novel [NiFe] hydrogenase subgroups.[33,34,56,62] Several genomic studies have also suggested that [FeFe] hydrogenases may be encoded by uncultivated DPANN archaea.[1,56,63] However, this remains debatable given that archaea seemingly lack the three maturation enzymes (HydEFG) required to synthesize the biologically unique catalytic H-cluster of the [FeFe] hydrogenase[64] and, to date, there is no evidence for archaeal [FeFe] hydrogenase activity.

Here, we systematically analyzed the diversity of [FeFe] hydrogenases in the domain Archaea by searching all publicly available species-representative isolate genomes and metagenome-assembled genomes (MAGs), as well as multiple MAGs generated for this study. We relied on two innovations to validate these findings. First, we used AlphaFold2-based structural modeling to test whether conserved gene clusters encode divergent hydrogenase complexes.[65,66] Critically, we then combined heterologous enzyme production with artificial maturation to confirm whether archaeal hydrogenases were catalytically active and displayed H-cluster spectroscopic signals.[67–69] Through this integrated approach, we demonstrate that archaea harbor active ultraminimal [FeFe] hydrogenases and identify unexpected complexes with distinct sequences, structures, and probable functions to previously described enzymes, including hybrid complexes with [NiFe] hydrogenases. These findings revise our understanding of the distribution and evolution of microbial H_2 metabolism, and have broad biological, chemical, and biotechnological ramifications.

RESULTS AND DISCUSSION

Structurally and genetically diverse [FeFe] hydrogenases are encoded by nine archaeal phyla

We searched for the gene encoding the catalytic subunit of [FeFe] hydrogenases (HydA) in the 2,339 archaeal species clusters of the Genome Taxonomy Database (GTDB) and our repository of unpublished archaeal MAGs. In total, 130 archaeal genomes (90 previously reported, 40 novel) encoded [FeFe] hydrogenases, spanning 9 phyla and 17 classes (Figure 1; Table S1). Except for some Asgard archaea, only one [FeFe] hydrogenase was present in each genome. The enzymes were verified and classified based on analysis of domain structure (Table S2; Figure S1), genetic organization (Table S2; Figure S2), maturases (Table S3; Figure 1), and primary phylogeny (detailed below). The enzymes fell into six distinct groups, namely the canonical groups A1 (n = 26), A3 (n = 44), and B (n = 12) and the groups E (n = 30), F (n = 21), and G (n = 3) defined in this work (Figure 1; Table S2). The archaeal group A1 and group E [FeFe] hydrogenases are putative fermentative enzymes encoded by three DPANN phyla (Iainarchaeota, Micrarchaeota, and Nanoarchaeota). With average sequences of just 363 and 286 residues, respectively (after excluding any truncated sequences), these enzymes are much smaller than the most minimal hydrogenase previously characterized (*Chlamydomonas reinhardtii* HydA1; 457 residues[64,70]) (Table S2). The most widespread hydrogenase, however, is the electron-bifurcating/confurcating

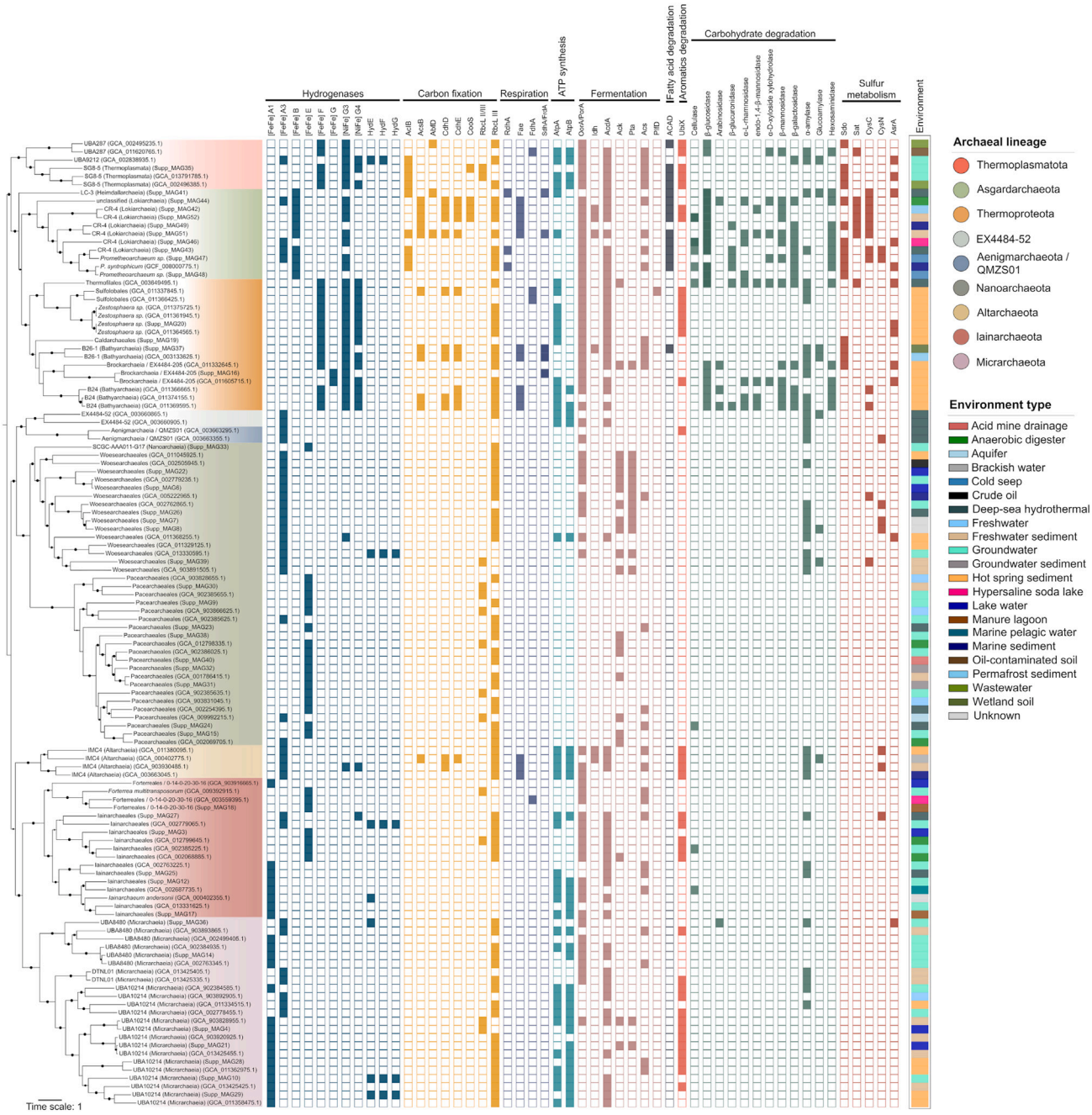


Figure 1. Phylogenetically and metabolically diverse archaea encode [FeFe] hydrogenases

The left portion of the figure shows a maximum-likelihood phylogenomic tree (model LG + F + G4) based on the concatenated 15 ribosomal marker proteins of archaeal genomes that encode [FeFe] hydrogenases. Results are shown for the 118 (out of 130) genomes that are at least 60% complete, less than 5% contaminated, and contain at least 75% of the 15 syntenic proteins. Branches are color coded, encoding according to the respective phylum. Black circles indicate bootstrap support values over 80%. The middle portion shows the presence of key metabolic genes (in at least one genome) involved in different metabolic processes. Carbon fixation: ATP-citrate lyase beta subunit (AclB), acetyl-CoA synthase beta subunit (AcsB), 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA-delta-isomerase (AbfD), carbon monoxide dehydrogenase/ acetyl-CoA synthase (CODH/ACS) complex subunit delta (CdhD), CODH/ACS complex subunit gamma (CdhE), anaerobic CODH catalytic subunit (CooS), type II/III ribulose-bisphosphate carboxylase (RbcL II/II), and type III ribulose-bisphosphate carboxylase (RbcL III); respiration: reductive dehalogenase (RdhA), formaldehyde activating enzyme (Fae), formate dehydrogenase subunit alpha (FdhA), and reversible succinate dehydrogenase and fumarate reductase flavoprotein (SdhA/FrdA); ATP synthesis: ATP synthase subunit alpha (AtpA) and ATP synthase subunit beta (AtpB); fermentation: 2-oxoacid:ferredoxin or pyruvate:ferredoxin oxidoreductase alpha subunit (OorA/PorA), L-lactate dehydrogenase

(legend continued on next page)

group A3 [FeFe] hydrogenase. This hydrogenase, together with its partner diaphorase (HydB) and thioredoxin (HydC) subunits, is encoded by at least six DPANN phyla, Thermoplasmatota (class E2), and some Asgard archaea (class Lokiarchaeia) (Figure 1).

Among cultured archaea, [FeFe] hydrogenases are only encoded in the Asgard archaeal enrichment cultures. We detected group B [FeFe] hydrogenases in the genomes of both "*Ca.* P. syntrophicum"[35] (Figure 1; Table S1) and our recently reported culture "*Candidatus* Lokiarchaeum ossiferum"[53] (Table S4). Although "*Ca.* P. syntrophicum" has previously been inferred to fermentatively produce H_2, this activity was assumed to originate from its group 3c [NiFe] hydrogenase and its [FeFe] hydrogenase was misannotated as an $F_{420}H_2$-dependent dehydrogenase subunit[35]; based on previously reported transcriptomes,[35] the [FeFe] hydrogenase is expressed at similarly high levels (309 reads per kilobase million [RPKM]) to the [NiFe] hydrogenase (333 RPKM), suggesting that it may contribute to observed H_2 production. We also conducted proteomic and transcriptomic analyses to gain insights into the metabolic capabilities of "*Ca.* L. ossiferum" (Table S4); this archaeon also synthesizes an [FeFe] hydrogenase at high levels (18.03 label-free quantification [LFQ] intensity [$\log_2$]), but its [NiFe] hydrogenase (24.00 LFQ intensity [$\log_2$]) is among the most abundant complexes in the cell and thus likely dominates H_2 production (Table S4). Nevertheless, the [FeFe] hydrogenase may contribute to H_2 production, given that these enzymes typically have higher activities than their [NiFe] counterparts and all biochemically characterized group 3c [NiFe] hydrogenases oxidize H_2 under cellular conditions.[1,71,72] Given the extremely slow growth rates and yields of both cultures,[35,53] we are currently unable to conduct deeper analyses of the differential roles of these enzymes *in vivo*. Diverse [FeFe] hydrogenases are also encoded by MAGs of several other Lokiarchaeia and Heimdallarchaeia (Figure 1; Table S1).

To better understand the structure and function of putative archaeal [FeFe] hydrogenases, we performed structural modeling using AlphaFold2. Archaeal group A1 and E [FeFe] hydrogenases are monomeric enzymes (HydA only) predicted to fold into a H-cluster domain with a solvent-exposed catalytic H-cluster (Figures 2A and 2B; Figures S3A and S3B). However, in contrast to bacterial [FeFe] hydrogenases, the loop structures of these enzymes are condensed and additional iron-sulfur clusters are absent (Figure S4). Each of the modeled enzymes contained the cysteine residues required to ligate the H-cluster, though differed in their proton-transferring residues (Figures S4A–S4C). Thus, despite their small size, these enzymes are theoretically capable of H_2 catalysis. The group B [FeFe] hydrogenase from "*Ca.* P. syntrophicum" (denoted Ps) modeled as a heterodimer between the HydA and HydC subunits encoded by the gene cluster; it contains two 2 × [4Fe-4S] ferredoxin-like domains, one that acts as an electron relay from the H-cluster and the other of unknown function separate from the main body of the protein (Figure 2C). Structural modeling also supported the notion that archaeal group A3 [FeFe] hydrogenases form trimeric electron-bifurcating complexes (HydABC) similar to those recently structurally characterized in fermentative and acetogenic bacteria[73,74] (Figure 2D; Figure S3D).

Archaeal [FeFe] hydrogenases are catalytically active and display H-cluster spectroscopic signals

We tested whether group A1, B, and E [FeFe] hydrogenases from archaea bind the catalytic H-cluster and produce H_2. To do so, we heterologously expressed five enzymes in *Escherichia coli* BL21(DE3), anaerobically matured them using the synthetic mimic $[2Fe]^{adt}$ ($[Fe_2(azadithiolate)(CO)_4(CN)_2]^{2-}$), and measured H_2 production of whole-cell lysates using gas chromatography as per established protocols (Table S5; Figure S5A).[67–69,75] H_2 production was clearly discernible relative to negative controls in enzymes matured from all three groups (Figure 3A; Table S5). The highest activity was observed from a Micrarchaeota group A1 enzyme (denoted Mu). Cell lysates containing Mu evolved H_2 at half the rate of the well-known [FeFe] hydrogenase HydA1 from the green alga *C. reinhardtii*, which was included in all assays as a positive control.[69] The homologous enzyme from the groundwater archaeon "*Ca.* Iainarchaeum andersonii"[76] (denoted Ia) also produced H_2, though at a 100-fold lower rate. We further observed substantial activity of the group B enzyme from the Asgard archaea (Ps)[35] and the group E enzyme from the groundwater archaeon "*Ca.* Forterrea multitransposorum"[77] (denoted Fm). We note that the observed H_2 production activities should not be considered as specific activities given potential variations in the efficiency of heterologous expression, folding, and maturation between the enzymes; this is illustrated by the differences in activity between the Mu and Ia hydrogenases despite their high degree of sequence homology. Nevertheless, these results validate that both DPANN and Asgard archaea encode functional [FeFe] hydrogenases and that the minimal group A1 and ultraminimal group E enzymes are both active.

The group E [FeFe] hydrogenase Fm was isolated as a representative example to provide more detailed insight into the properties of these ultra-minimalistic enzymes. Following purification under strictly anaerobic conditions and semi-enzymatic reconstitution of the iron-sulfur clusters, the iron content of the enzyme was determined to be 4.2 ± 0.4 per protein (Figures S5B and S5C), in agreement with the structural modeling indicating that this enzyme contains a single [4Fe4S] cluster (Figure 2). The reconstituted Fm hydrogenase was subsequently incubated with $[2Fe]^{adt}$. Successful H-cluster assembly was verified through attenuated total reflection Fourier transformed infrared (ATR-FTIR) spectroscopy, given that sharp cofactor bands in the expected CO/CN ligand band region of the FTIR spectra were

(Idh), ADP-forming acetyl-CoA synthetase (AcdA), acetate kinase (Ack), phosphate acetyltransferase (Pta), acetyl-CoA synthetase (Acs), and formate C-acetyltransferase (PflD); fatty acid degradation: acyl-CoA dehydrogenase (ACAD); aromatics degradation: flavin prenyltransferase (UbiX); sulfur metabolism: sulfur dioxygenase (Sdo), sulfate adenylyltransferase (Sat), adenylylsulfate kinase (CysC), sulfate adenylyltransferase subunit 1 (CysN), and anaerobic sulfite reductase subunit A (AsrA). The right portion shows the diverse environments from where the archaeal genomes were retrieved. Note that the phylum QMZS01 was classified as Aenigmatarchaeota in GTDB R06-RS207, while Thermoproteota class EX4484-205 was proposed as Brockarchaeia.
See also Figures S1–S3 and S6.

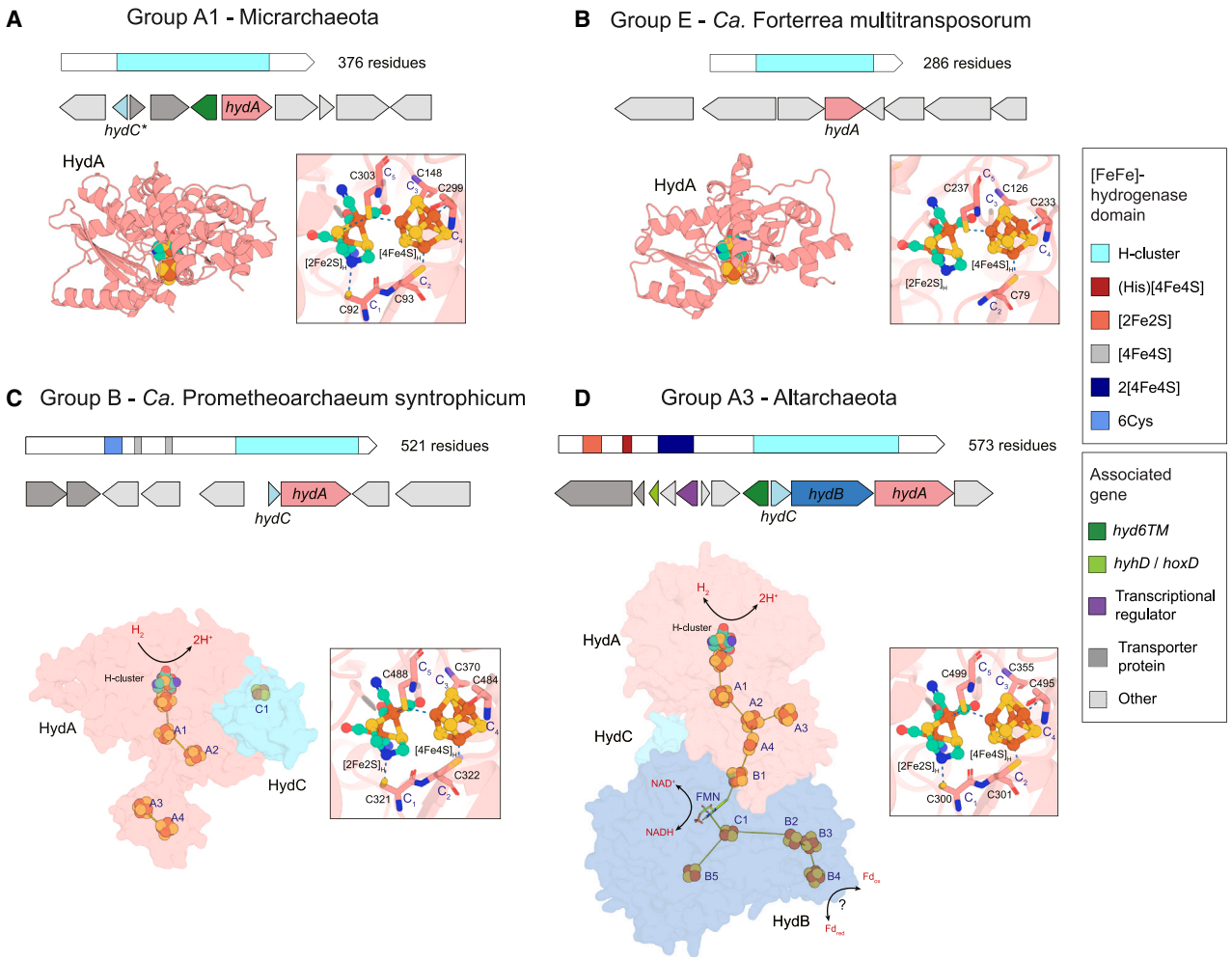


Figure 2. Archaea encode genetically and structurally diverse [FeFe] hydrogenases

Catalytic domain structure, genetic organization, and AlphaFold2-based structural modeling of representative [FeFe] hydrogenases encoded in archaeal genomes.

(A) Group A1 [FeFe] hydrogenase from UBA95 sp002499405 (Micrarchaeota).

(B) Group E [FeFe] hydrogenase from "*Ca.* Forterrea multitransposorum".

(C) Group B [FeFe] hydrogenase complex from "*Ca.* Prometheoarchaeum syntrophicum."

(D) Group A3 [FeFe] hydrogenase complex from DSAL01 sp011380095 (Altarchaeota).

For each panel, the catalytic domain (H-cluster), iron-sulfur binding motifs, and amino acid sequence length are shown at the top. Genes encoding hydrogenase structural subunits are shown in their genetic context beneath, labeled and colored consistent with the corresponding subunit in the structural models. Predicted cofactors are positioned based on the structures of homologous proteins. A zoomed view of the H-cluster and conserved coordinating cysteine residues (C_1 to C_5) is shown for each group. For group B and A3 enzymes, FeS clusters within plausible electron transfer distance are connected by dashed lines. *hydA*, [FeFe] hydrogenase; *hydB*, diaphorase; *hydC*, thioredoxin; *hydD*, *nuoG*-like conduit protein; *hyd6TM*, uncharacterized 4- to 6-helix transmembrane protein associated with group A [FeFe] hydrogenases; (His)[4Fe4S], $(Cys)_3$His-ligated [4Fe4S] cluster binding domain; [2Fe2S], [2Fe2S] cluster binding domain; [4Fe4S], [4Fe4S] cluster binding domain; 2[4Fe4S], bacterial ferredoxin-like 2[4Fe4S] cluster binding domain; 6Cys, putative iron-sulfur cluster binding domain. * HydC protein in group A1 gene cluster was not predicted to form a complex with HydA. Surface structures are used for the multisubunit group B and A3 [FeFe] hydrogenases, with ribbon diagram versions provided in Figure S3.

See also Figures S1–S4.

readily observed (Figure 3B; Figure S5D). Our spectroscopic analysis suggested that the enzyme was isolated in an inhibited state. However, photochemical reduction resulted in the transition to catalytically active states (Figure 3B; Figure S5E) reminiscent of the oxidized active ready states (H_{ox} and $H_{ox}H$[78]) and a further reduced state (H_{red}'[79,80]).

The catalytic properties of Fm were studied by protein film electrochemistry (PFE). Cyclic voltammetry traces of the enzyme recorded under a H_2 atmosphere showed the typical bidirectional catalytic behavior commonly associated with [FeFe] hydrogenases[81] (Figure 3C). A comparison of the reducing and oxidizing currents observed at high driving force (±300 mV vs.

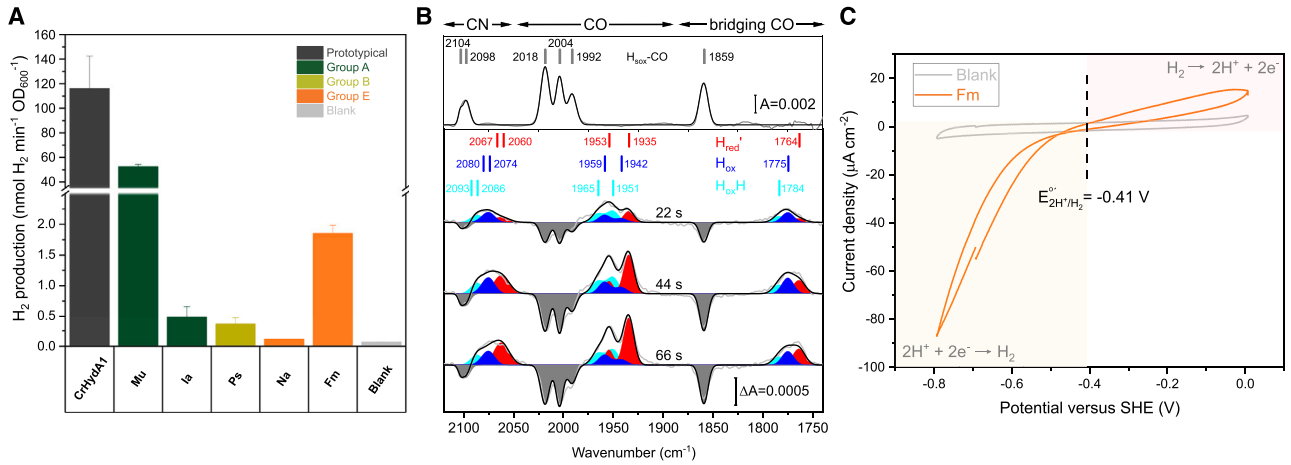


Figure 3. Three classes of [FeFe] hydrogenases encoded by archaea are catalytically active

(A) H_2 gas production monitored from cell lysates in *E. coli* BL21(DE3) cells expressing group A1, B, and E [FeFe] hydrogenases from archaea. All cell lysates, including the blank, were activated by addition of $[2Fe]^{adt}$. H_2 was measured by gas chromatography (GC) after addition of methyl viologen and dithionite to activated cell lysates, set to pH 6.8 with 100 mM KPi buffer. Activities are normalized for number of cells used (nmol H_2 min^{-1} $OD_{600}{}^{-1}$) and error bars reflect standard deviation from biological triplicates. The strain expressing prototypical *Cr*HydA1 was used as a positive control while "blank" represents the same strain but containing an empty vector.

(B) FTIR spectra of the group E [FeFe] hydrogenase from "*Ca.* Forterrea multitransposorum" (Fm) after heterologous expression, semisynthetic maturation with $[2Fe]^{adt}$, and purification. The absorbance spectrum (top) indicates a CO-inhibited di-ferrous H-cluster state (H_{sox}-CO). The difference spectra (bottom) illustrates the transitions of Fm into catalytically active states through photoreduction (illumination after the addition of eosin Y as a photosensitizer and triethanolamine as a sacrificial electron donor). During illumination, bands associated with the highly oxidized CO-inhibited state decreased (gray bands), while new bands reflecting reduced and catalytically active H-cluster states appear, assigned to $H_{ox}H$ (cyan), H_{ox} (blue), and H_{red} (red) (spectra arranged chronologically from top to bottom).

(C) Cyclic voltammetry traces of immobilized Fm (orange) with H_2 oxidation current densities at high potentials and H^+ reduction currents at low potentials. The $2H^+/H_2$ redox couple potential is indicated with a dashed line ($E^{o'}{}_{2H^+/H_2}$). Scan direction is indicated by black arrows. The blank trace (gray) represents the electrode without an immobilized enzyme film. The experiments were performed on two independent films for each enzyme at pH 7.0 (5 mM MES, 5 mM CHES, 5 mM HEPES, 5 mM TAPS, 5 mM sodium acetate [NaOAc], 0.1 M Na_2SO_4) and under 1 atm H_2.

See also Figures S5.

reversible hydrogen electrode [RHE]) indicated that the enzyme is clearly biased toward H^+ reduction catalysis relative to H_2 gas oxidation. This is in contrast to its closest characterized homolog, the group C [FeFe] hydrogenase from the bacterium *Thermotoga maritima* (28% sequence identity), which showed a clear preference for H_2 oxidation.[41] Collectively, these results show that archaeal monomeric [FeFe] hydrogenases bind a redox-active H-cluster in a very well-defined environment and suggest that they mediate fermentative H_2 production.

[FeFe] and [NiFe] hydrogenases associate into complexes in uncultivated archaea

Remarkably, the group F [FeFe] hydrogenases appear to form complexes with [NiFe] hydrogenases. 21 genomes encoded these complexes through five gene clusters, including from the classes Bathyarchaeia, Brockarchaeia, Thermoplasmata, Thermoproteia, and Lokiarchaeia (Figure 1; Table S2). Their defining feature is the fusion of a C-terminal [FeFe] hydrogenase catalytic domain (HydA) with an N-terminal domain homologous to the group 3 [NiFe] hydrogenase small subunit (HyhS) (Figure 4A; Figures S3E and S3F). Four other genes are contiguous with this fusion: the large subunit of the group 3 [NiFe] hydrogenase (HyhL), the diaphorase (HydB) and thioredoxin (HydC) subunits of the electron-bifurcating/confurcating group A3 [FeFe] hydrogenase, and a conduit subunit containing four iron-sulfur clusters (herein HydD) (Figure S2). The thioredoxin, diaphorase, and conduit subunits are, respectively, homologous to NuoE (subunit E), NuoF (subunit F), and NuoG (subunit G), which together form the NADH dehydrogenase module of complex I.[82] All four genes are also present in the recently identified and structurally characterized electron-bifurcating [NiFe] hydrogenase from *Acetomicrobium mobile*[14]; however, the archaeal complex is distinct, given it contains a true [FeFe] hydrogenase catalytic subunit with an H-cluster domain that is fused to [NiFe] hydrogenase small subunit. Of the archaeal [FeFe] hydrogenases, the group F enzymes show the most conserved genetic organization in archaea, presumably due to their association with [NiFe] hydrogenases (Figure S2). The genetically and phylogenetically distinct group G [FeFe] hydrogenases, exclusive to the Brockarchaeia genus JAAOZO01, are also likely to form hybrid complexes (Figure S2; Table S2).

To confirm the catalytic activity of the group F [FeFe] hydrogenase, we recombinantly expressed and artificially matured two HydA-HyhS fusion proteins from the candidate lineage Thermoplasmata SG8-5. One of these enzymes rapidly produced H_2 over the time course (Th1), with relative activities of 5% compared with the *C. reinhardtii* enzyme (Figure 4D; Table S5). In contrast, the other enzyme (Th2) exhibited only low levels of

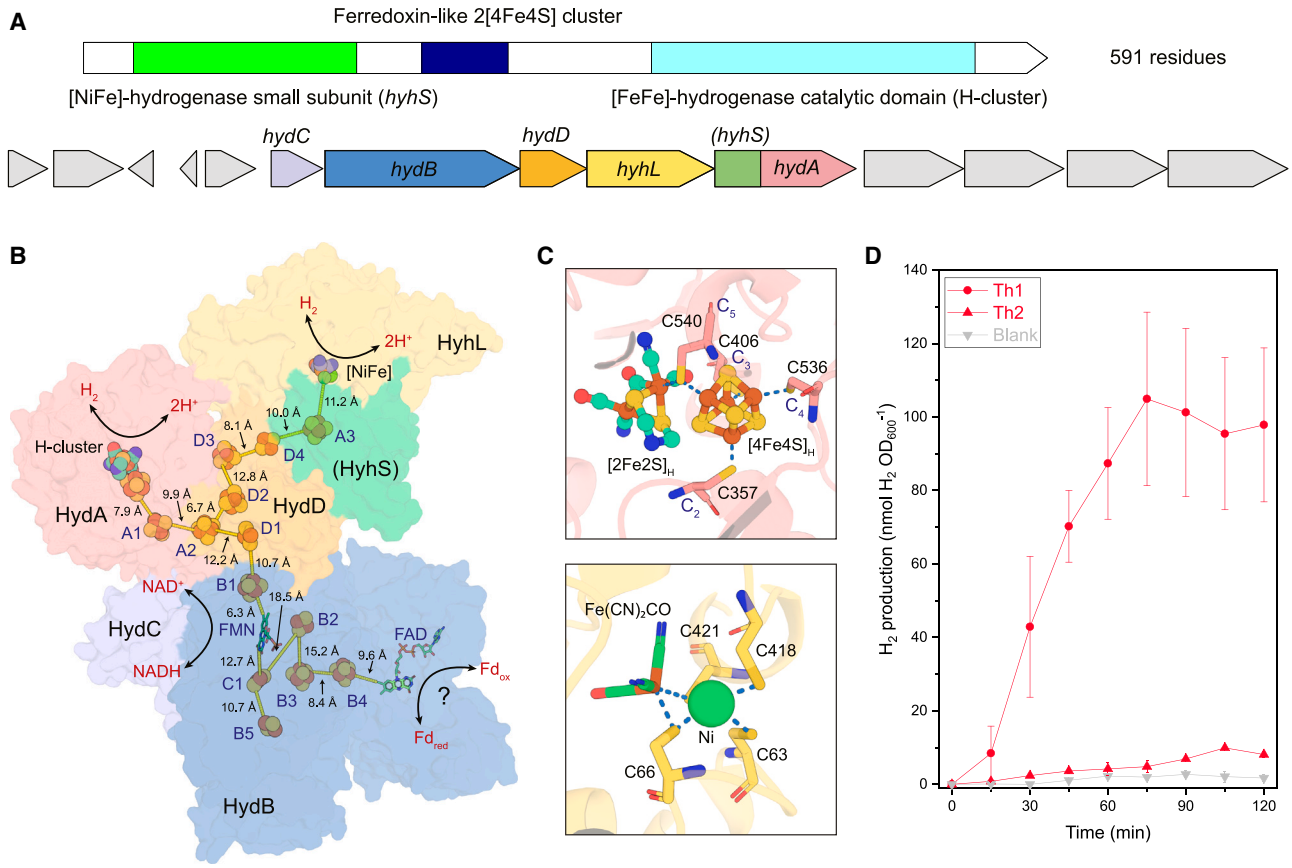


Figure 4. [FeFe] and [NiFe] hydrogenases encoded by archaea are predicted to form unique complexes

(A) Catalytic domain structure and predicted operon encoding a putative complex of a group F [FeFe] hydrogenase and group 3 [NiFe] hydrogenase in Thermoplasmatota UBA147 sp002496385. *hydA*, [FeFe] hydrogenase catalytic domain (fused with *hyhS*); *hydB*, diaphorase; *hydC*, thioredoxin; *hydD*, *nuoG*-like conduit protein; *hyhL*, group 3 [NiFe] hydrogenase catalytic subunit; *hyhS*, group 3 [NiFe] hydrogenase small/iron-sulfur domain (fused with *hydA*).

(B) Predicted surface structure, cofactor composition, and electron flow through four potential arms in the hybrid hydrogenase complex. [FeS] clusters are numbered and labeled according to their subunit of origin (e.g., A1, A2, and A3 originate from the HydA subunit).

(C) Atomic structure of the predicted [FeFe]- and [NiFe]-hydrogenase active sites in the hybrid enzyme. Distances between catalytic cluster and coordinating residues of less than 2.5 Å are shown as blue dotted lines.

(D) H_2 gas production monitored from cell lysates in *E. coli* BL21(DE3) cells expressing the Th1 and Th2 [FeFe] hydrogenases from archaea. The cell lysates were activated by addition of $[2Fe]^{adt}$. H_2 levels were measured every 15 min for 2 h by gas chromatography after addition of methyl viologen and dithionite to activated cell lysates, set to pH 6.8 with 100 mM KPi buffer. Activities are normalized for number of cells used (nmol H_2 OD_{600}^{-1}) and error bars reflect standard deviations from two biological triplicates. Blank represents the same strain but contains an empty vector.

See also Figures S1–S4 and S7.

activity (Figure 4D). The proteins proved challenging to purify, preventing a detailed *in vitro* characterization. It is likely that these enzymes would display even higher activity if assembled into a complex with the other subunits, though this would be exceptionally challenging to achieve given that the archaea encoding them are uncultivated and [NiFe] hydrogenases are recalcitrant to heterologous production. These measurements nevertheless confirm that these are *bona fide* hydrogenases and that their activities are attributable to the H-cluster.

Finally, we performed AlphaFold2 modeling to infer whether the [FeFe] and [NiFe] hydrogenases associate (Figures 4B and 4C). When modeled alone, the hybrid subunit was predicted to contain a [FeFe] hydrogenase catalytic domain and a [NiFe] hydrogenase iron-sulfur cluster domain separated by a long flexible linker (Figure S3F). Conserved cysteine ligands required to ligate both the H-cluster of the [FeFe] hydrogenase and the three electron-relaying [4Fe4S] clusters of the [NiFe] hydrogenase component were both observed (Figure 4C; Figures S4A–S4C). However, modeling of all five genetically contiguous subunits (HydA-HyhS, HydBCD, HyhL) suggests that they form a stable electron-bifurcating/confurcating complex (ΔG = −97.06 to −219.7 kJ mol^{-1}; PISA analysis) that associates through multiple hydrogen bonds and salt bridges (Figure S3I). The structural model suggests that the complex receives electrons through the [FeFe] hydrogenase arm or the [NiFe] hydrogenase via a series of iron-sulfur clusters to a possible electron-converging [4Fe4S] cluster on the hybrid subunit, or vice versa. Thereafter, electrons are predicted to be simultaneously transferred to the

high-potential NAD^+ at the HydC subunit and an undetermined low-potential acceptor (likely ferredoxin) at the glutamate synthase (GltA) domain of the HydB subunit (Figure 4B). These observations are remarkable, given that [NiFe] and [FeFe] hydrogenases are not known to associate. Moreover, they suggest surprising modularity of the [NiFe] hydrogenase small subunit, given it has evidently co-evolved with both [NiFe] and [FeFe] hydrogenases.

[FeFe] hydrogenases enable fermentation and electron bifurcation in diverse archaea

We sought to understand the role of the various [FeFe] hydrogenases in the metabolism of archaea. To do so, we annotated the high-quality archaeal genomes for genes associated with major energy conservation and carbon acquisition processes. All [FeFe] hydrogenase-encoding archaea are predicted to be obligate anaerobes, given they lacked terminal oxidases. This is consistent with the retrieval of the genomes from typically anoxic ecosystems, especially groundwater, anaerobic digesters, and sediments from hot springs, hydrothermal vents, and freshwater (Figure 1; Table S1). Based on the retrieved genomes, most DPANN archaea are likely to be symbiotic obligate fermenters dependent on host-derived organic compounds; consistently, they often encoded genes for the degradation and fermentation of carbohydrates (primarily starch) and aromatic compounds, but generally lacked respiratory reductases or carbon fixation pathways (Figure 1). The other archaeal phyla were predicted to be capable of a wider range of metabolic strategies, in line with their larger genome sizes (Figure S1), including β-oxidation, anaerobic respiration, and carbon fixation to varying extents (Figure 1). Notably, some Asgard archaea encoded fumarate reductases (Frd), reductive dehalogenases (Rdh), and anaerobic sulfite reductases (Asr), with the latter enzyme also encoded in certain MAGs from four other phyla. Several lineages were also predicted to be capable of autotrophy through the Wood-Ljungdahl (Lokiarchaeia, Thermoproteota) or reverse tricarboxylic acid (Lokiarchaeia, Heimdallarchaeia, Thermoplasmatota) pathways (Figure 1; Table S1). Most of the MAGs also encoded ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) lineages known to function in nucleoside salvage.[83]

The group A, B, and E [FeFe] hydrogenases likely facilitate cofactor regeneration during organic carbon fermentation in diverse archaea. Half of the archaea encode 2-oxoacid-ferredoxin oxidoreductases, such as pyruvate-ferredoxin oxidoreductases that couple the oxidation of the end product of glycolysis (pyruvate) to the reduction of ferredoxin, and most can gain ATP by converting the derived acetyl-coenzyme A (CoA) to the end product acetate via the acetyl-CoA synthetase or acetate kinase reaction (Figure 1). As supported by the structural modeling (Figure 2), the monomeric group A1, B, and E hydrogenases are predicted to couple ferredoxin reoxidation to H_2 production. Also consistent with this role, 2-oxoacid-ferredoxin oxidoreductase genes are frequently adjacent to [FeFe] hydrogenase genes in Nanoarchaeota and Micrarchaeota MAGs (Figure S2). By contrast, the trimeric group A3 [FeFe] hydrogenases are predicted to simultaneously reoxidize ferredoxin (reduced primarily by pyruvate-ferredoxin oxidoreductase) and NADH (e.g., reduced during glycolysis) (Figure 2), in line with their bacterial counterparts.[11,73,74] Congruently, the electron-bifurcating/confurcating group A3 [FeFe] hydrogenases are associated with those DPANN archaea harboring relatively complex carbohydrate degradation pathways (i.e., Woesearchaeles, Altarchaeota, some Iainarchaeota) through which both NAD^+ and ferredoxin will be reduced (Figure 1). It is also notable that many DPANN archaea encode the minimal group A1 and ultraminimal group E [FeFe] hydrogenases as their sole H_2-metabolizing enzymes (Table S1); in conjunction with the simpler maturation pathways of [FeFe] hydrogenases compared with [NiFe] hydrogenases, these genome-reduced microorganisms (average completeness-normalized genome size: 1.2 Mbp) have minimized the genetic and cellular costs of metabolizing H_2. Further studies are needed to determine what biochemical features differentiate the monomeric group A1, B, and E [FeFe] hydrogenases and whether they have distinct physiological roles. It cannot be ruled out that the group B enzymes instead consume H_2 to support anaerobic respiration or carbon fixation in Asgard archaea; however, this seems unlikely given their reported H_2-evolving activities (Figure 3), structural features (Figure 2), and the hydrogenogenic lifestyle of "*Ca.* P. syntrophicum."[35]

The physiological role of the hybrid hydrogenases is unclear. These enzymes are exclusively encoded by more metabolically versatile archaea, including those with the capacity for carbon fixation, β-oxidation, and energy-conversion using group 4 [NiFe] hydrogenases (Figure 1). The predicted structures suggest that that these enzymes contribute to electron bifurcation by transferring electrons from H_2 to both NAD^+ and likely ferredoxin, or vice versa (Figure 3). However, it is peculiar that a single complex contains two seemingly redundant hydrogenase modules. A potential explanation is that one of the hydrogenase modules might act on a substrate other than H_2, especially given group 3 [NiFe] hydrogenases can serve as sulfhydrogenases *in vitro* (i.e., mediating reduction of elemental sulfur to hydrogen sulfide).[84] A further rationale is that the two hydrogenase modules may differ in their affinities and/or oxygen tolerance, enabling efficient H_2 oxidation across a wide range of environmental conditions. However, perhaps the strongest possibility is that the complexes act as redox valves, transferring electrons either from NADH and reduced ferredoxin to a H_2-evolving [FeFe] hydrogenase when reductant accumulates or from a H_2-consuming [NiFe] hydrogenase to NAD^+ and oxidized ferredoxin otherwise; this is consistent with previous observations that electron-bifurcating hydrogenases act as H_2-producing redox valves during heterotrophic growth of bacteria in response to variations in substrate availability and redox state.[13] Other enzyme complexes regulating opposite reactions have recently been reported, namely between glutamate synthase (GltAB) and glutamate dehydrogenase (GudB),[85] with the GltA domain of these enzymes also shared in the hybrid hydrogenase complex.

[FeFe] hydrogenases have been acquired by archaea on multiple occasions and have an ancient association with [NiFe] hydrogenases

We investigated the evolutionary history of [FeFe] hydrogenases through phylogenetic analysis of its catalytic subunit (Figure 5) and three maturases (Figure S6). In agreement with their

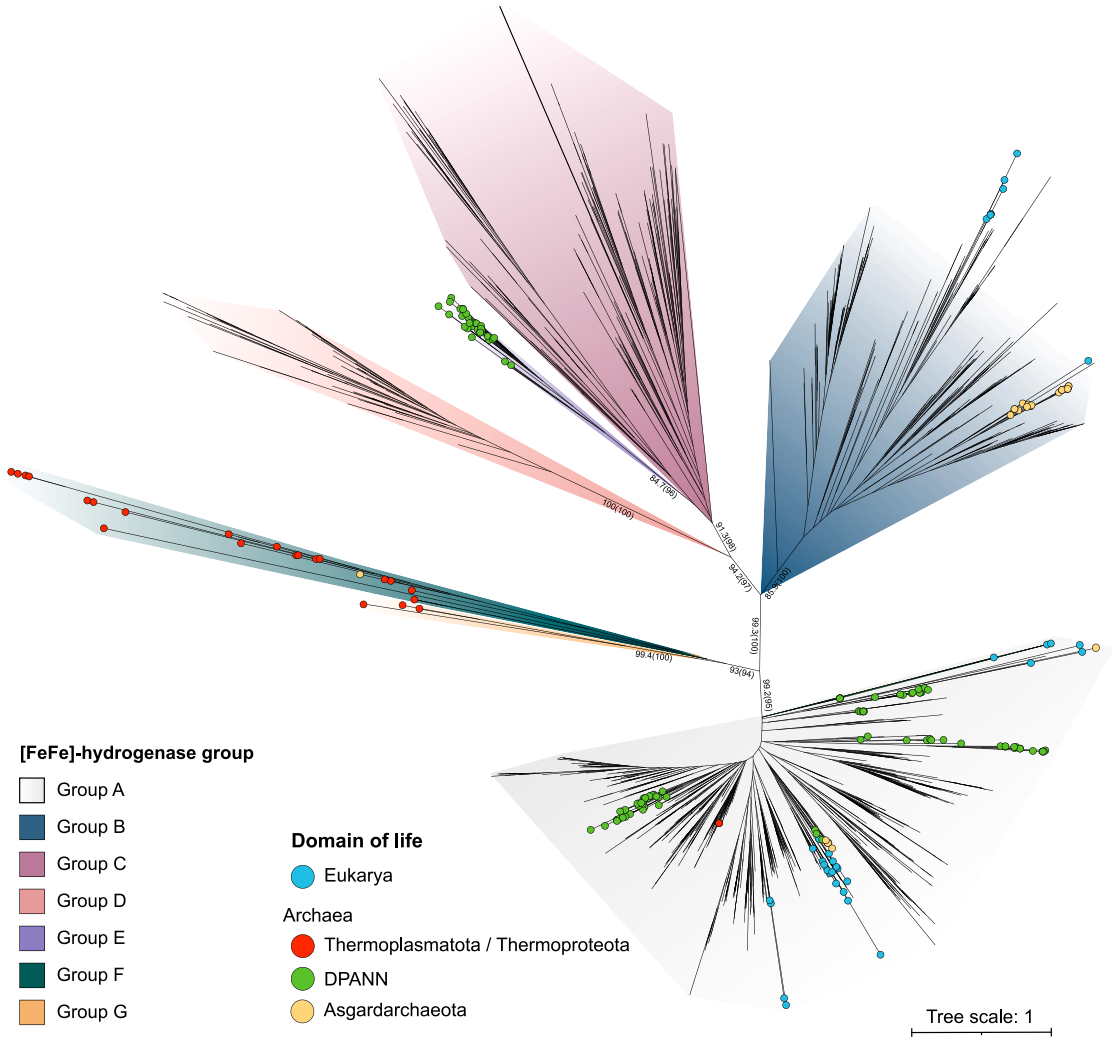


Figure 5. [FeFe] hydrogenases are diverse and ancient in archaea

An unrooted maximum-likelihood phylogenetic tree of the catalytic subunit (HydA) of [FeFe] hydrogenases and the hybrid hydrogenases. The tree was constructed based on 3,677 amino acid sequences using the LG + C20 + R + F model. The numbers at the branches indicate the aLRT (approximate likelihood ratio test) and ultrafast bootstrap (within bracket) support values, each with 1,000 replicates. The scale bar corresponds to the expected number of substitutions per site. Colored circles at the tip indicate sequences from eukaryotes and major archaeal groups. All other sequences are from bacteria.
See also Figures S1–S3.

classification into known groups, the archaeal group A1, A3, and B [FeFe] hydrogenases clustered with various bacterial and eukaryotic homologs; the group A enzymes form at least six radiations, suggesting that they were laterally acquired from bacterial enzymes over several independent events, whereas the archaeal group B and E enzymes each formed a monophyletic clade (Figure 5). Based on phylogenies constructed using the best supported model (Figure 5) and their structural simplicity (Figure 2B), the ultraminimal fermentative group E enzymes of archaea form a sister clade to the multidomain sensory group C hydrogenases of bacteria.[1,41,86,87] We also re-examined the putative origin of [FeFe] hydrogenases in eukaryotes in light of the syntrophy hypotheses for eukaryogenesis,[31–36] given our finding that [FeFe] hydrogenases are present both in Asgard archaea and unicellular eukaryotes. Our data do not support the hypothesis that all eukaryotic [FeFe] hydrogenases were vertically acquired from an Asgard archaeal ancestor. The eukaryotic enzymes cluster into at least five disparate clades, spanning both groups A and B, suggesting multiple horizontal acquisitions (Figure 5). In addition, [FeFe] hydrogenases are absent from the Asgard genomes most closely related to the archaeal ancestor of eukaryotes (Figure 1) and the alphaproteobacterial genomes most related to mitochondria,[1,88,89] though this view could change with the addition of new genomes. Nevertheless, archaeal and eukaryotic [FeFe] hydrogenases clustered together in three of these lineages (including Lokiarchaeia with *Tritrichomonas*), suggesting that some eukaryotes and archaea may have laterally exchanged [FeFe] hydrogenase genes during their diversification (Figure 5).

We also studied the distribution and phylogeny of the three maturases that synthesize the [FeFe] hydrogenase cofactor, HydE, HydF, and HydG. Of the 130 archaeal genomes encoding [FeFe] hydrogenases, just five encode a full set of maturases and three others encode an incomplete set (Table S3). Though genome incompleteness means that the co-occurrence of hydrogenases and maturases will be underestimated, this does not explain the absence of maturases in most genomes. Indeed, maturase genes are even absent in the genome of the cultured hydrogenogenic archaeon "*Ca.* P. syntrophicum"[35] and the closed genome of "*Ca.* F. multitransposorum" from which an active [FeFe] hydrogenase was heterologously produced (Figure 3). In turn, it is likely that archaea synthesize [FeFe] hydrogenases through an alternative pathway. The existence of an alternative pathway is consistent with reports that various eukaryotes lacking all (*Giardia*, *Entamoeba*[90,91]) or some (*Mastigamoeba*[92]) maturases still make catalytically active H_2-producing hydrogenases, as well as recent reports of a cytosolic [FeFe] hydrogenase in *Trichomonas vaginalis*[93] and the reported heterologous production of active [FeFe] hydrogenases in *Synechocystis* cells lacking maturases.[94] In line with findings for the structural subunits, phylogenetic analysis of HydE, HydF, and HydG suggests that archaea acquired maturases on several occasions (Figures S6). Several archaeal maturases are closely related to those that we recently identified in Chlamydiae and eukaryotes.[89]

Phylogenetic analyses also suggest an ancient origin of the hybrid hydrogenases. In phylogenetic trees of the [FeFe] hydrogenase catalytic subunit, sequences of group F and G hydrogenases together formed a monophyletic branch distant from the group A and group B–E superclades, suggesting an early divergence of these archaeal enzymes (Figure 5). These sequences formed long branches, given their divergence from other hydrogenases (~25% sequence identity to the model hydrogenase *C. reinhardtii* HydA1). The discovery of these divergent archaeal hydrogenases raises the question of whether [FeFe] hydrogenase first evolved in archaea or bacteria; however, the absence of non-hydrogenase homologs ancestral to [FeFe] hydrogenases precludes the confident rooting of the tree needed to make strong inferences. In phylogenetic trees of [NiFe] hydrogenases, the catalytic (large) subunits of the fusion proteins formed multiple clusters (Figure S7A). In contrast, the iron-sulfur (small) domain fused to the group F [FeFe] hydrogenase and the iron-sulfur subunit downstream of the group G [FeFe] hydrogenase each form distinct monophyletic subgroups within the group 3 [NiFe] hydrogenases (Figure S7B). Thus, as dictated by the gene fusion, the iron-sulfur domain of hybrid complexes more strongly co-evolved with the [FeFe]- than the [NiFe]-hydrogenase catalytic subunits. We propose that the components of the hybrid hydrogenases are formally recognized as distinct lineages, namely the group F and G [FeFe] hydrogenases and group 3f and 3g [NiFe] hydrogenases, given their distinct phylogenies, structures, and potential physiological roles.

Limitations of the study

Although our combinatorial approach shows that [FeFe] hydrogenases from archaea are active, this study does not directly observe their activity in their native hosts. Almost all archaea encoding these enzymes are uncultured, with the exception of two extremely slow-growing, low-yield enrichment cultures of Asgard archaea, and thus no genetic or biochemical systems are currently available to study them in native hosts. Though heterologously expressed archaeal hydrogenases incorporated a synthetic catalytic H-cluster and often displayed high rates of catalytic activity, it cannot be ruled out that some archaea natively produce a distinct but related catalytic cluster for H_2 catalysis through an unknown pathway, especially given that most archaea lack the three conventional [FeFe] hydrogenase maturases. Given its reliance on heterologous expression and semisynthetic maturation, this study cannot determine the native activity rates of these hydrogenases, especially for those enzymes predicted to form multisubunit complexes. In addition, while genome-wide metabolic analysis and structural modeling are useful to infer the possible functions of [FeFe] hydrogenases in the host cell, extensive experiments in host cells are necessary to confirm their physiological roles and interactions.

Conclusions

Archaea have evolved remarkably disparate ways to use [FeFe] hydrogenases to adapt to anoxic environments. On one hand, DPANN archaea have evolved ultraminimal enzymes to efficiently dispose of reductant derived from carbohydrate fermentation. Conversely, lineages such as Brockarchaeia and Lokiarchaeia use distinct hybrids of [NiFe] and [FeFe] hydrogenases—among the most complex hydrogenases described to date—to support their diverse redox biology. These discoveries expand the [FeFe] hydrogenases from four to seven groups, each with distinct phylogenies, structures, and functions. In addition to increasing our understanding of archaeal biology, these findings also redefine our understanding of hydrogen metabolism and hydrogenase biochemistry by revealing: (1) [FeFe] hydrogenases from all three domains of life are active, (2) the two dominant hydrogenase classes have co-evolved, and (3) a size minimum for hydrogenases. The ultraminimal hydrogenases also provide ideal templates both to understand enzymatic H_2 catalysis, for example, determinants of rate, directionality, affinity, and oxygen sensitivity, as well as flexible scaffolds for directed evolution of efficient H_2-converting biocatalysts. More broadly, our approach also emphasizes the potential of combining genome-resolved metagenomics with accurate protein structure prediction and heterologous production studies to discover new enzymes and functions in uncultured microorganisms.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Strains, media and growth conditions
- METHOD DETAILS
  - Genome sources

- Metagenomic assembly and binning
- Identification of archaeal [FeFe] hydrogenase genes
- Analysis of [FeFe] hydrogenase domain architecture
- Analysis of [FeFe] hydrogenase genetic organization
- Identification of archaeal [FeFe] hydrogenase maturases
- '*Ca.* Lokiarchaeum ossiferum' transcriptomic analysis
- '*Ca.* Lokiarchaeum ossiferum' proteomic analysis
- AlphaFold2 structural modelling
- Protein expression and characterization
- Hydrogenase semisynthetic maturation
- Hydrogenase purification
- Activation of purified hydrogenase
- Protein film electrochemistry
- ATR-FTIR spectroscopy
- Genome-wide metabolic annotations
- Genome-level phylogenetic analysis
- Gene-level phylogenetic analysis

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.05.032.

ACKNOWLEDGMENTS

This work was supported by two National Health & Medical Research Council Emerging Leader Fellowships (APP1178715 to C.G.; APP1197376 to R.G.), two Australian Research Council Discovery Project grants (DP200103074 and DP230103080, both awarded to C.G. and R.G.), the Swedish Energy Agency (STEM 48574-1 to G.B.), the National Science Foundation Partnerships for International Research and Education grant (OISE-2230766 to C.G. and J.F.B.), the Olle Engkvists stiftelse (220-0226 to G.B. and M. Senger), the Swedish Research Council (Vetenskapsrådet Starting Grant 2020-05071 to C.W.S.), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ERC Starting grant 947317 to A.S. and 101078476 to C.W.S.), the Bill & Melinda Gates Foundation (to J.F.B. and L.E.V.A.), a University of California Dissertation-Year Fellowship (to L.E.V.A.), an Australian Government Research Training Stipend Scholarship (to P.M.L.), a Monash International Tuition Scholarship (to P.M.L.), and a Monash FMNHS Early Career Postdoctoral Fellowship (ECPF23-1113137961 to P.M.L.). Work on "*Ca.* L. ossiferum" was funded by the European Research Council (AdG TACKLE: 695192, StG EVOLPHYSIOL: 803768) and the Austrian Science Fund (FWF: Z437). We thank Alexander L. Jaffe, Yuki Amano, Madalena Alves, José Leal, Ricardo Leite, Nelson Simões, and Duarte Toubarro for providing access to genomes. Ping Huang is gratefully acknowledged for support during spectroscopy data acquisition and analysis, and Tristan Wagner, Gerrit Schut, Masaru Nobu, and C.S. Raman are thanked for helpful comments. We thank the MonARCH HPC Cluster and the M3 MASSIVE HPC facility for providing computation platforms. Proteomics analyses were performed by the Mass Spectrometry Facility at Max Perutz Labs, Vienna, Austria, using the VBCF instrument pool.

AUTHOR CONTRIBUTIONS

C.G. conceived and oversaw this study. C.G. discovered the archaeal and hybrid [FeFe] hydrogenases. J.F.B., L.E.V.A., S.M., and J.W.-R. contributed original metagenomic datasets and metagenome-assembled genomes. P.M.L., M.M., C.G., L.E.V.A., J.M., J.S., and R.L. conducted homology-based searches. P.M.L., H.L., C.G., and G.B. analyzed domain and genetic organization. R.G. performed structural modeling. T.R.-O., R.I.P.-T., and C.S. conducted transcriptomic and proteomic analysis. P.R.C., H.L., and G.B. conducted heterologous expression and activity assays. P.R.C., M.A.K., M. Senger, and G.B. performed spectroscopic analysis. P.R.C. and M.A.K. performed enzyme isolation and analysis. P.R.C. and G.B. performed protein film electrochemistry analysis. L.E.V.A., P.M.L., C.G., M. Schoelmerich, and J.F.B. conducted genome-wide metabolic analysis. L.E.V.A., A.S., C.G., C.W.S., J.F.B., and P.M.L. performed phylogenetic analysis. C.G., P.M.L., P.R.C., G.B., and R.G. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

J.F.B. is a co-founder of Metagenomi. A patent on this discovery and application of ultraminimal hydrogenases was submitted.

Received: June 17, 2023
Revised: April 12, 2024
Accepted: May 16, 2024
Published: June 11, 2024

REFERENCES

1. Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M., and Morales, S.E. (2016). Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. ISME J. *10*, 761–777. https://doi.org/10.1038/ismej.2015.153.
2. Schwartz, E., Fritsch, J., and Friedrich, B. (2013). H_2-metabolizing prokaryotes. In The Prokaryotes: Prokaryotic Physiology and Biochemistry, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson, eds. (Springer), pp. 119–199. https://doi.org/10.1007/978-3-642-30141-4_65.
3. Peters, J.W., Schut, G.J., Boyd, E.S., Mulder, D.W., Shepard, E.M., Broderick, J.B., King, P.W., and Adams, M.W.W. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochim. Biophys. Acta *1853*, 1350–1369. https://doi.org/10.1016/j.bbamcr.2014.11.021.
4. Volbeda, A., Charon, M.H., Piras, C., Hatchikian, E.C., Frey, M., and Fontecilla-Camps, J.C. (1995). Crystal structure of the nickel–iron hydrogenase from *Desulfovibrio gigas*. Nature *373*, 580–587. https://doi.org/10.1038/373580a0.
5. Peters, J.W., Lanzilotta, W.N., Lemon, B.J., and Seefeldt, L.C. (1998). X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom resolution. Science *282*, 1853–1858. https://doi.org/10.1126/science.282.5395.1853.
6. Shima, S., Pilak, O., Vogt, S., Schick, M., Stagni, M.S., Meyer-Klaucke, W., Warkentin, E., Thauer, R.K., and Ermler, U. (2008). The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. Science *321*, 572–575. https://doi.org/10.1126/science.1158978.
7. Lubitz, W., Ogata, H., Rüdiger, O., and Reijerse, E. (2014). Hydrogenases. Chem. Rev. *114*, 4081–4148. https://doi.org/10.1021/cr4005814.
8. Vignais, P.M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. Chem. Rev. *107*, 4206–4272. https://doi.org/10.1021/cr050196r.
9. Pinske, C. (2019). Bioenergetic aspects of archaeal and bacterial hydrogen metabolism. In Advances in Microbial Physiology, R.K. Poole, ed. (Academic Press), pp. 487–514. https://doi.org/10.1016/bs.ampbs.2019.02.005.
10. Leung, P.M., Grinter, R., Tudor-Matthew, E., Lingford, J.P., Jimenez, L., Milton, M., Hanchapola, I., Tanuwidjaya, E., Kropp, A., Peach, H.A., et al. (2024). Trace gas oxidation sustains energy needs of a thermophilic archaeon at suboptimal temperatures. Nat. Commun. *15*, 3219. https://doi.org/10.1038/s41467-024-47324-2.
11. Schut, G.J., and Adams, M.W.W. (2009). The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J. Bacteriol. *191*, 4451–4457. https://doi.org/10.1128/jb.01582-08.
12. Schuchmann, K., and Müller, V. (2012). A bacterial electron-bifurcating hydrogenase. J. Biol. Chem. *287*, 31165–31171. https://doi.org/10.1074/jbc.M112.395038.

13. Wang, S., Huang, H., Kahnt, J., and Thauer, R.K. (2013). A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*. J. Bacteriol. *195*, 1267–1275. https://doi.org/10.1128/jb.02158-12.

14. Feng, X., Schut, G.J., Haja, D.K., Adams, M.W.W., and Li, H. (2022). Structure and electron transfer pathways of an electron-bifurcating NiFe-hydrogenase. Sci. Adv. *8*, eabm7546. https://doi.org/10.1126/sciadv.abm7546.

15. Buckel, W., and Thauer, R.K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Evol. Asp. Biochim. Biophys. Acta *1827*, 94–113. https://doi.org/10.1016/j.bbabio.2012.07.002.

16. Schuchmann, K., Chowdhury, N.P., and Müller, V. (2018). Complex multimeric [FeFe] hydrogenases: biochemistry, physiology and new opportunities for the hydrogen economy. Front. Microbiol. *9*, 2911. https://doi.org/10.3389/fmicb.2018.02911.

17. Schut, G.J., Haja, D.K., Feng, X., Poole, F.L., Li, H., and Adams, M.W.W. (2022). An abundant and diverse new family of electron bifurcating enzymes with a non-canonical catalytic mechanism. Front. Microbiol. *13*, 946711. https://doi.org/10.3389/fmicb.2022.946711.

18. Søndergaard, D., Pedersen, C.N.S., and Greening, C. (2016). HydDB: a web tool for hydrogenase classification and analysis. Sci. Rep. *6*, 34212. https://doi.org/10.1038/srep34212.

19. Piché-Choquette, S., and Constant, P. (2019). Molecular hydrogen, a neglected key driver of soil biogeochemical processes. Appl. Environ. Microbiol. *85*, e02418-18. https://doi.org/10.1128/AEM.02418-18.

20. Constant, P., Poissant, L., and Villemur, R. (2009). Tropospheric H_2 budget and the response of its soil uptake under the changing environment. Sci. Total Environ. *407*, 1809–1823. https://doi.org/10.1016/j.scitotenv.2008.10.064.

21. Greening, C., Islam, Z.F., and Bay, S.K. (2022). Hydrogen is a major lifeline for aerobic bacteria. Trends Microbiol. *30*, 330–337. https://doi.org/10.1016/j.tim.2021.08.004.

22. Morita, R.Y. (1999). Is H_2 the universal energy source for long-term survival? Microb. Ecol. *38*, 307–320. https://doi.org/10.1016/j.tim.2021.08.004.

23. Benoit, S.L., Maier, R.J., Sawers, R.G., and Greening, C. (2020). Molecular hydrogen metabolism: a widespread trait of pathogenic bacteria and protists. Microbiol. Mol. Biol. Rev. *84*, e00092-19. https://doi.org/10.1128/MMBR.00092-19.

24. Carbonero, F., Benefiel, A.C., and Gaskins, H.R. (2012). Contributions of the microbial hydrogen economy to colonic homeostasis. Nat. Rev. Gastroenterol. Hepatol. *9*, 504–518. https://doi.org/10.1038/nrgastro.2012.85.

25. Lee, H.-S., Vermaas, W.F.J., and Rittmann, B.E. (2010). Biological hydrogen production: prospects and challenges. Trends Biotechnol. *28*, 262–271. https://doi.org/10.1016/j.tibtech.2010.01.007.

26. Cracknell, J.A., Vincent, K.A., and Armstrong, F.A. (2008). Enzymes as working or inspirational electrocatalysts for fuel cells and electrolysis. Chem. Rev. *108*, 2439–2461. https://doi.org/10.1021/cr0680639.

27. Evans, R.M., Siritanaratkul, B., Megarity, C.F., Pandey, K., Esterle, T.F., Badiani, S., and Armstrong, F.A. (2019). The value of enzymes in solar fuels research – efficient electrocatalysts through evolution. Chem. Soc. Rev. *48*, 2039–2052. https://doi.org/10.1039/C8CS00546J.

28. Kleinhaus, J.T., Wittkamp, F., Yadav, S., Siegmund, D., and Apfel, U.-P. (2021). [FeFe]-hydrogenases: maturation and reactivity of enzymatic systems and overview of biomimetic models. Chem. Soc. Rev. *50*, 1668–1784. https://doi.org/10.1039/D0CS01089H.

29. Lane, N., Allen, J.F., and Martin, W. (2010). How did LUCA make a living? Chemiosmosis in the origin of life. BioEssays *32*, 271–280. https://doi.org/10.1002/bies.200900131.

30. Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., and Martin, W.F. (2016). The physiology and habitat of the last universal common ancestor. Nat. Microbiol. *1*, 16116. https://doi.org/10.1038/nmicrobiol.2016.116.

31. Martin, W., and Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. Nature *392*, 37–41. https://doi.org/10.1038/32096.

32. Moreira, D., and López-García, P. (1998). Symbiosis between methanogenic archaea and δ-Proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. *47*, 517–530. https://doi.org/10.1007/PL00006408.

33. Sousa, F.L., Neukirchen, S., Allen, J.F., Lane, N., and Martin, W.F. (2016). Lokiarchaeon is hydrogen dependent. Nat. Microbiol. *1*, 16034. https://doi.org/10.1038/nmicrobiol.2016.34.

34. Spang, A., Stairs, C.W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E.F., Greening, C., Baker, B.J., and Ettema, T.J.G. (2019). Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. Nat. Microbiol. *4*, 1138–1148. https://doi.org/10.1038/s41564-019-0406-9.

35. Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., et al. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. Nature *577*, 519–525. https://doi.org/10.1038/s41586-019-1916-6.

36. López-García, P., and Moreira, D. (2020). The Syntrophy hypothesis for the origin of eukaryotes revisited. Nat. Microbiol. *5*, 655–667. https://doi.org/10.1038/s41564-020-0710-4.

37. Land, H., Senger, M., Berggren, G., and Stripp, S.T. (2020). Current state of [FeFe]-hydrogenase research: biodiversity and spectroscopic investigations. ACS Catal. *10*, 7069–7086. https://doi.org/10.1021/acscatal.0c01614.

38. Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. Microbiology (Reading) *156*, 1575–1588. https://doi.org/10.1099/mic.0.032771-0.

39. Greening, C., Ahmed, F.H., Mohamed, A.E., Lee, B.M., Pandey, G., Warden, A.C., Scott, C., Oakeshott, J.G., Taylor, M.C., and Jackson, C.J. (2016). Physiology, biochemistry, and applications of F_{420}- and F_o-dependent redox reactions. Microbiol. Mol. Biol. Rev. *80*, 451–493. https://doi.org/10.1128/mmbr.00070-15.

40. Dietrich, H.M., Righetto, R.D., Kumar, A., Wietrzynski, W., Trischler, R., Schuller, S.K., Wagner, J., Schwarz, F.M., Engel, B.D., Müller, V., et al. (2022). Membrane-anchored HDCR nanowires drive hydrogen-powered CO_2 fixation. Nature *607*, 823–830. https://doi.org/10.1038/s41586-022-04971-z.

41. Chongdar, N., Birrell, J.A., Pawlak, K., Sommer, C., Reijerse, E.J., Rüdiger, O., Lubitz, W., and Ogata, H. (2018). Unique spectroscopic properties of the H-cluster in a putative sensory [FeFe] hydrogenase. J. Am. Chem. Soc. *140*, 1057–1068. https://doi.org/10.1021/jacs.7b11287.

42. Land, H., Sekretareva, A., Huang, P., Redman, H.J., Németh, B., Polidori, N., Mészáros, L.S., Senger, M., Stripp, S.T., and Berggren, G. (2020). Characterization of a putative sensory [FeFe]-hydrogenase provides new insight into the role of the active site architecture. Chem. Sci. *11*, 12789–12801. https://doi.org/10.1039/D0SC03319G.

43. Morra, S. (2022). Fantastic [FeFe]-hydrogenases and where to find them. Front. Microbiol. *13*, 853626. https://doi.org/10.3389/fmicb.2022.853626.

44. Ortiz, M., Leung, P.M., Shelley, G., Jirapanjawat, T., Nauer, P.A., Van Goethem, M.W.V., Bay, S.K., Islam, Z.F., Jordaan, K., Vikram, S., et al. (2021). Multiple energy sources and metabolic strategies sustain microbial diversity in Antarctic desert soils. Proc. Natl. Acad. Sci. USA *118*, e2025322118. https://doi.org/10.1073/pnas.2025322118.

45. Zirngibl, C., Hedderich, R., and Thauer, R.K. (1990). N^5,N^{10}-Methylenetetrahydromethanopterin dehydrogenase from *Methanobacterium thermoautotrophicum* has hydrogenase activity. FEBS Lett. *261*, 112–116. https://doi.org/10.1016/0014-5793(90)80649-4.

46. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature *428*, 37–43. https://doi.org/10.1038/nature02340.

47. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431–437. https://doi.org/10.1038/nature12352.

48. Castelle, C.J., and Banfield, J.F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell *172*, 1181–1197. https://doi.org/10.1016/j.cell.2018.02.016.

49. Spang, A., Caceres, E.F., and Ettema, T.J.G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science *357*, eaaf3883. https://doi.org/10.1126/science.aaf3883.

50. Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R.V., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521*, 173–179. https://doi.org/10.1038/nature14447.

51. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358. https://doi.org/10.1038/nature21031.

52. Liu, Y., Makarova, K.S., Huang, W.-C., Wolf, Y.I., Nikolskaya, A.N., Zhang, X., Cai, M., Zhang, C.-J., Xu, W., Luo, Z., et al. (2021). Expanded diversity of Asgard archaea and their relationships with eukaryotes. Nature *593*, 553–557. https://doi.org/10.1038/s41586-021-03494-3.

53. Rodrigues-Oliveira, T., Wollweber, F., Ponce-Toledo, R.I., Xu, J., Rittmann, S.K.R., Klingl, A., Pilhofer, M., and Schleper, C. (2023). Actin cytoskeleton and complex cell architecture in an Asgard archaeon. Nature *613*, 332–339. https://doi.org/10.1038/s41586-022-05550-y.

54. Eme, L., Tamarit, D., Caceres, E.F., Stairs, C.W., De Anda, V., Schön, M.E., Seitz, K.W., Dombrowski, N., Lewis, W.H., Homa, F., et al. (2023). Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature *618*, 992–999. https://doi.org/10.1038/s41586-023-06186-2.

55. Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr. Biol. *25*, 690–701. https://doi.org/10.1016/j.cub.2015.01.014.

56. Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H., and Banfield, J.F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat. Rev. Microbiol. *16*, 629–645. https://doi.org/10.1038/s41579-018-0076-2.

57. Dombrowski, N., Lee, J.-H., Williams, T.A., Offre, P., and Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. FEMS Microbiol. Lett. *366*, fnz008. https://doi.org/10.1093/femsle/fnz008.

58. Laso-Pérez, R., Wegener, G., Knittel, K., Widdel, F., Harding, K.J., Krukenberg, V., Meier, D.V., Richter, M., Tegetmeyer, H.E., Riedel, D., et al. (2016). Thermophilic archaea activate butane via alkyl-coenzyme M formation. Nature *539*, 396–401. https://doi.org/10.1038/nature20152.

59. Chen, S.-C., Musat, N., Lechtenfeld, O.J., Paschke, H., Schmidt, M., Said, N., Popp, D., Calabrese, F., Stryhanyuk, H., Jaekel, U., et al. (2019). Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. Nature *568*, 108–111. https://doi.org/10.1038/s41586-019-1063-0.

60. Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019). An evolving view of methane metabolism in the Archaea. Nat. Rev. Microbiol. *17*, 219–232. https://doi.org/10.1038/s41579-018-0136-7.

61. Dong, X., Greening, C., Rattray, J.E., Chakraborty, A., Chuvochina, M., Mayumi, D., Dolfing, J., Li, C., Brooks, J.M., Bernard, B.B., et al. (2019). Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. Nat. Commun. *10*, 1816. https://doi.org/10.1038/s41467-019-09747-0.

62. Borrel, G., Adam, P.S., McKay, L.J., Chen, L.X., Sierra-García, I.N., Sieber, C.M.K., Letourneur, Q., Ghozlane, A., Andersen, G.L., Li, W.J., et al. (2019). Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. Nat. Microbiol. *4*, 603–613. https://doi.org/10.1038/s41564-019-0363-3.

63. Huang, W.-C., Liu, Y., Zhang, X., Zhang, C.-J., Zou, D., Zheng, S., Xu, W., Luo, Z., Liu, F., and Li, M. (2021). Comparative genomic analysis reveals metabolic flexibility of Woesearchaeota. Nat. Commun. *12*, 5281. https://doi.org/10.1038/s41467-021-25565-9.

64. Mulder, D.W., Boyd, E.S., Sarma, R., Lange, R.K., Endrizzi, J.A., Broderick, J.B., and Peters, J.W. (2010). Stepwise [FeFe]-hydrogenase H-cluster assembly revealed in the structure of $HydA^{\Delta EFG}$. Nature *465*, 248–251. https://doi.org/10.1038/nature08993.

65. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

66. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. https://doi.org/10.1101/2021.10.04.463034.

67. Berggren, G., Adamska, A., Lambertz, C., Simmons, T.R., Esselborn, J., Atta, M., Gambarelli, S., Mouesca, J.M., Reijerse, E., Lubitz, W., et al. (2013). Biomimetic assembly and activation of [FeFe]-hydrogenases. Nature *499*, 66–69. https://doi.org/10.1038/nature12239.

68. Khanna, N., Esmieu, C., Mészáros, L.S., Lindblad, P., and Berggren, G. (2017). *In vivo* activation of an [FeFe] hydrogenase using synthetic cofactors. Energy Environ. Sci. *10*, 1563–1567. https://doi.org/10.1039/C7EE00135E.

69. Land, H., Ceccaldi, P., Mészáros, L.S., Lorenzi, M., Redman, H.J., Senger, M., Stripp, S.T., and Berggren, G. (2019). Discovery of novel [FeFe]-hydrogenases for biocatalytic H_2-production. Chem. Sci. *10*, 9941–9948. https://doi.org/10.1039/C9SC03717A.

70. Winkler, M., Esselborn, J., and Happe, T. (2013). Molecular basis of [FeFe]-hydrogenase function: an insight into the complex interplay between protein and catalytic cofactor. Biochim. Biophys. Acta *1827*, 974–985. https://doi.org/10.1016/j.bbabio.2013.03.004.

71. Costa, K.C., Wong, P.M., Wang, T., Lie, T.J., Dodsworth, J.A., Swanson, I., Burn, J.A., Hackett, M., and Leigh, J.A. (2010). Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. Proc. Natl. Acad. Sci. USA *107*, 11050–11055. https://doi.org/10.1073/pnas.1003653107.

72. Kaster, A.-K., Moll, J., Parey, K., and Thauer, R.K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. Proc. Natl. Acad. Sci. USA *108*, 2981–2986. https://doi.org/10.1073/pnas.1016761108.

73. Furlan, C., Chongdar, N., Gupta, P., Lubitz, W., Ogata, H., Blaza, J.N., and Birrell, J.A. (2022). Structural insight on the mechanism of an electron-bifurcating [FeFe] hydrogenase. eLife *11*, e79361. https://doi.org/10.7554/eLife.79361.

74. Katsyv, A., Kumar, A., Saura, P., Pöverlein, M.C., Freibert, S.A., T Stripp, S., Jain, S., Gamiz-Hernandez, A.P., Kaila, V.R.I., Müller, V., et al. (2023). Molecular basis of the electron bifurcation mechanism in the [FeFe]-hydrogenase complex HydABC. J. Am. Chem. Soc. *145*, 5696–5709. https://doi.org/10.1021/jacs.2c11683.

75. Esselborn, J., Lambertz, C., Adamska-Venkates, A., Simmons, T., Berggren, G., Noth, J., Siebel, J., Hemschemeier, A., Artero, V., Reijerse, E., et al. (2013). Spontaneous activation of [FeFe]-hydrogenases by an

inorganic [2Fe] active site mimic. Nat. Chem. Biol. *9*, 607–609. https://doi.org/10.1038/nchembio.1311.

76. Youssef, N.H., Rinke, C., Stepanauskas, R., Farag, I., Woyke, T., and El-shahed, M.S. (2015). Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. ISME J. *9*, 447–460. https://doi.org/10.1038/ismej.2014.141.

77. Probst, A.J., and Banfield, J.F. (2018). Homologous recombination and transposon propagation shape the population structure of an organism from the deep subsurface with minimal metabolism. Genome Biol. Evol. *10*, 1115–1119. https://doi.org/10.1093/gbe/evy067.

78. Senger, M., Mebs, S., Duan, J., Shulenina, O., Laun, K., Kertess, L., Wittkamp, F., Apfel, U.-P., Happe, T., Winkler, M., et al. (2018). Protonation/reduction dynamics at the [4Fe–4S] cluster of the hydrogen-forming cofactor in [FeFe]-hydrogenases. Phys. Chem. Chem. Phys. *20*, 3128–3140. https://doi.org/10.1039/C7CP04757F.

79. Sommer, C., Adamska-Venkatesh, A., Pawlak, K., Birrell, J.A., Rüdiger, O., Reijerse, E.J., and Lubitz, W. (2017). Proton coupled electronic rearrangement within the H-cluster as an essential step in the catalytic cycle of [FeFe] hydrogenases. J. Am. Chem. Soc. *139*, 1440–1443. https://doi.org/10.1021/jacs.6b12636.

80. Laun, K., Baranova, I., Duan, J., Kertess, L., Wittkamp, F., Apfel, U.-P., Happe, T., Senger, M., and Stripp, S.T. (2021). Site-selective protonation of the one-electron reduced cofactor in [FeFe]-hydrogenase. Dalton Trans. *50*, 3641–3650. https://doi.org/10.1039/D1DT00110H.

81. Vincent, K.A., Parkin, A., and Armstrong, F.A. (2007). Investigating and exploiting the electrocatalytic properties of hydrogenases. Chem. Rev. *107*, 4366–4413. https://doi.org/10.1021/cr050191u.

82. Sazanov, L.A. (2015). A giant molecular proton pump: structure and mechanism of respiratory complex I. Nat. Rev. Mol. Cell Biol. *16*, 375–388. https://doi.org/10.1038/nrm3997.

83. Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science *315*, 1003–1006. https://doi.org/10.1126/science.1135999.

84. Ma, K., Schicho, R.N., Kelly, R.M., and Adams, M.W. (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. https://doi.org/10.1073/pnas.90.11.5341.

85. Jayaraman, V., Lee, D.J., Elad, N., Vimer, S., Sharon, M., Fraser, J.S., and Tawfik, D.S. (2022). A counter-enzyme complex regulates glutamate metabolism in *Bacillus subtilis*. Nat. Chem. Biol. *18*, 161–170. https://doi.org/10.1038/s41589-021-00919-y.

86. Greening, C., Geier, R., Wang, C., Woods, L.C., Morales, S.E., McDonald, M.J., Rushton-Green, R., Morgan, X.C., Koike, S., Leahy, S.C., et al. (2019). Diverse hydrogen production and consumption pathways influence methane production in ruminants. ISME J. *13*, 2617–2632. https://doi.org/10.1038/s41396-019-0464-2.

87. Zheng, Y., Kahnt, J., Kwon, I.H., Mackie, R.I., and Thauer, R.K. (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. https://doi.org/10.1128/jb.02070-14.

88. Stairs, C.W., Táborský, P., Salomaki, E.D., Kolisko, M., Pánek, T., Eme, L., Hradilová, M., Vlček, Č., Jerlström-Hultqvist, J., Roger, A.J., et al. (2021). Anaeramoebae are a divergent lineage of eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes. Curr. Biol. *31*, 5605–5612.e5. https://doi.org/10.1016/j.cub.2021.10.010.

89. Stairs, C.W., Dharamshi, J.E., Tamarit, D., Eme, L., Jørgensen, S.L., Spang, A., and Ettema, T.J.G. (2020). Chlamydial contribution to anaerobic metabolism during eukaryotic evolution. Sci. Adv. *6*, eabb7258. https://doi.org/10.1126/sciadv.abb7258.

90. Nixon, J.E.J., Field, J., McArthur, A.G., Sogin, M.L., Yarlett, N., Loftus, B.J., and Samuelson, J. (2003). Iron-dependent hydrogenases of *Entamoeba histolytica* and *Giardia lamblia*: activity of the recombinant entamoebic enzyme and evidence for lateral gene transfer. Biol. Bull. *204*, 1–9. https://doi.org/10.2307/1543490.

91. Lloyd, D., Ralphs, J.R., and Harris, J.C. (2002). *Giardia intestinalis*, a eukaryote without hydrogenosomes, produces hydrogen. Microbiology (Reading) *148*, 727–733. https://doi.org/10.1099/00221287-148-3-727.

92. Nývltová, E., Šuták, R., Harant, K., Šedinová, M., Hrdý, I., Pačes, J., Vlček, Č., and Tachezy, J. (2013). NIF-type iron-sulfur cluster assembly system is duplicated and distributed in the mitochondria and cytosol of *Mastigamoeba balamuthi*. Proc. Natl. Acad. Sci. USA *110*, 7371–7376. https://doi.org/10.1073/pnas.1219590110.

93. Smutná, T., Dohnálková, A., Sutak, R., Narayanasamy, R.K., Tachezy, J., and Hrdý, I. (2022). A cytosolic ferredoxin-independent hydrogenase possibly mediates hydrogen uptake in *Trichomonas vaginalis*. Curr. Biol. *32*, 124–135.e5. https://doi.org/10.1016/j.cub.2021.10.050.

94. Berto, P., D'Adamo, S., Bergantino, E., Vallese, F., Giacometti, G.M., and Costantini, P. (2011). The cyanobacterium *Synechocystis* sp. PCC 6803 is able to express an active [FeFe]-hydrogenase without additional maturation proteins. Biochem. Biophys. Res. Commun. *405*, 678–683. https://doi.org/10.1016/j.bbrc.2011.01.095.

95. Li, H., and Rauchfuss, T.B. (2002). Iron carbonyl sulfides, formaldehyde, and amines condense to give the proposed azadithiolate cofactor of the Fe-only hydrogenases. J. Am. Chem. Soc. *124*, 726–727. https://doi.org/10.1021/ja016964n.

96. Zaffaroni, R., Rauchfuss, T.B., Gray, D.L., De Gioia, L., and Zampella, G. (2012). Terminal vs bridging hydrides of diiron dithiolates: protonation of Fe_2(dithiolate)$(CO)_2(PMe_3)_4$. J. Am. Chem. Soc. *134*, 19260–19269. https://doi.org/10.1021/ja3094394.

97. Cloirec, A.L., Davies, S.C., Evans, D.J., Hughes, D.L., Pickett, C.J., Best, S.P., and Borg, S. (1999). A di-iron dithiolate possessing structural elements of the carbonyl/cyanide sub-site of the H-centre of Fe-only hydrogenase. Chem. Commun. *22*, 2285–2286. https://doi.org/10.1039/A906391I.

98. Schmidt, M., Contakes, S.M., and Rauchfuss, T.B. (1999). First generation analogues of the binuclear site in the Fe-only hydrogenases: $Fe_2(\mu\text{-}SR)_2(CO)_4(CN)_2^{2-}$. J. Am. Chem. Soc. *121*, 9736–9737. https://doi.org/10.1021/ja9924187.

99. Lyon, E.J., Georgakaki, I.P., Reibenspies, J.H., and Darensbourg, M.Y. (1999). Carbon monoxide and cyanide ligands in a classical organometallic complex model for Fe-only hydrogenase. Angew. Chem. Int. Ed. Engl. *38*, 3178–3180. https://doi.org/10.1002/(SICI)1521-3773(19991102)38:21<3178::AID-ANIE3178>3.0.CO;2-4.

100. Loiseau, L., Ollagnier-de Choudens, S., Lascoux, D., Forest, E., Fontecave, M., and Barras, F. (2005). Analysis of the heteromeric CsdA-CsdE cysteine desulfurase, assisting Fe-S cluster biogenesis in *Escherichia coli*. J. Biol. Chem. *280*, 26760–26769. https://doi.org/10.1074/jbc.M504067200.

101. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A.A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. *36*, 996–1004. https://doi.org/10.1038/nbt.4229.

102. Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. *42*, D581–D591. https://doi.org/10.1093/nar/gkt1099.

103. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596. https://doi.org/10.1038/s41586-021-03828-1.

104. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). Blast+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

105. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

106. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

107. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. https://doi.org/10.1101/gr.186072.114.

108. Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. *27*, 135–145. https://doi.org/10.1002/pro.3290.

109. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

110. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

111. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638. https://doi.org/10.1093/molbev/msw046.

112. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

113. Wheeler, T.J., and Eddy, S.R. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics *29*, 2487–2489. https://doi.org/10.1093/bioinformatics/btt403.

114. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

115. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296. https://doi.org/10.1093/nar/gkab301.

116. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics *36*, 2251–2252. https://doi.org/10.1093/bioinformatics/btz859.

117. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

118. Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., and Anantharaman, K. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome *10*, 33. https://doi.org/10.1186/s40168-021-01213-8.

119. Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. Nat. Commun. *9*, 2542. https://doi.org/10.1038/s41467-018-04964-5.

120. Hollenstein, D.M., Maurer-Granofszky, M., Reiter, W., Anrather, D., Gossenreiter, T., Babic, R., Hartl, N., Kraft, C., and Hartl, M. (2023). Chemical acetylation of ligands and two-step digestion protocol for reducing codigestion in affinity purification–mass spectrometry. J. Proteome Res. *22*, 3383–3391. https://doi.org/10.1021/acs.jproteome.3c00424.

121. Mistry, J., Bateman, A., and Finn, R.D. (2007). Predicting active site residue annotations in the Pfam database. BMC Bioinformatics *8*, 298. https://doi.org/10.1186/1471-2105-8-298.

122. Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119. https://doi.org/10.1186/1471-2105-11-119.

123. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics *26*, 1608–1615. https://doi.org/10.1093/bioinformatics/btq249.

124. Krissinel, E. (2015). Stock-based detection of protein oligomeric states in jsPISA. Nucleic Acids Res. *43*, W314–W319. https://doi.org/10.1093/nar/gkv314.

125. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

126. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

127. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

128. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

129. Senger, M., Kernmayr, T., Lorenzi, M., Redman, H.J., and Berggren, G. (2022). Hydride state accumulation in native [FeFe]-hydrogenase with the physiological reductant H_2 supports its catalytic relevance. Chem. Commun. (Camb) *58*, 7184–7187. https://doi.org/10.1039/D2CC00671E.

130. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. *50*, D785–D794. https://doi.org/10.1093/nar/gkab776.

131. Castelle, C.J., Méheust, R., Jaffe, A.L., Seitz, K., Gong, X., Baker, B.J., and Banfield, J.F. (2021). Protein family content uncovers lineage relationships and bacterial pathway maintenance mechanisms in DPANN archaea. Front. Microbiol. *12*, 660052. https://doi.org/10.3389/fmicb.2021.660052.

132. Jaffe, A.L., Bardot, C., Le Jeune, A.-H., Liu, J., Colombet, J., Perrière, F., Billard, H., Castelle, C.J., Lehours, A.-C., and Banfield, J.F. (2023). Variable impact of geochemical gradients on the functional potential of bacteria, archaea, and phages from the permanently stratified Lac Pavin. Microbiome *11*, 14. https://doi.org/10.1186/s40168-022-01416-7.

133. Probst, A.J., Castelle, C.J., Singh, A., Brown, C.T., Anantharaman, K., Sharon, I., Hug, L.A., Burstein, D., Emerson, J.B., Thomas, B.C., et al. (2017). Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO_2 concentrations. Environ. Microbiol. *19*, 459–474. https://doi.org/10.1111/1462-2920.13362.

134. De Anda, V., Chen, L.-X., Dombrowski, N., Hua, Z.-S., Jiang, H.-C., Banfield, J.F., Li, W.-J., and Baker, B.J. (2021). Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. Nat. Commun. *12*, 2404. https://doi.org/10.1038/s41467-021-22736-6.

135. He, C., Keren, R., Whittaker, M.L., Farag, I.F., Doudna, J.A., Cate, J.H.D., and Banfield, J.F. (2021). Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. Nat. Microbiol. *6*, 354–365. https://doi.org/10.1038/s41564-020-00840-5.

136. Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., et al.

(2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. *7*, 13219. https://doi.org/10.1038/ncomms13219.

137. Al-Shayeb, B., Schoelmerich, M.C., West-Roberts, J., Valentin-Alvarado, L.E., Sachdeva, R., Mullen, S., Crits-Christoph, A., Wilkins, M.J., Williams, K.H., Doudna, J.A., et al. (2022). Borgs are giant genetic elements with potential to expand metabolic capacity. Nature *610*, 731–736. https://doi.org/10.1038/s41586-022-05256-1.

138. Valentin-Alvarado, L.E., Fakra, S.C., Probst, A.J., Giska, J.R., Jaffe, A.L., Oltrogge, L.M., West-Roberts, J., Rowland, J., Manga, M., Savage, D.F., et al. (2024). Autotrophic biofilms sustained by deeply-sourced groundwater host diverse CPR bacteria implicated in sulfur and hydrogen metabolism. Microbiome *12*, 15. https://doi.org/10.1101/2022.11.17.516901.

139. Méheust, R., Castelle, C.J., Matheus Carnevali, P.B., Farag, I.F., He, C., Chen, L.-X., Amano, Y., Hug, L.A., and Banfield, J.F. (2020). Groundwater Elusimicrobia are metabolically diverse compared to gut microbiome Elusimicrobia and some have a novel nitrogenase paralog. ISME J. *14*, 2907–2922. https://doi.org/10.1038/s41396-020-0716-1.

140. Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics *28*, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174.

141. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359. https://doi.org/10.7717/peerj.7359.

142. Nissen, J.N., Johansen, J., Allesøe, R.L., Sønderby, C.K., Armenteros, J.J.A., Grønbech, C.H., Jensen, L.J., Nielsen, H.B., Petersen, T.N., Winther, O., et al. (2021). Improved metagenome binning and assembly using deep variational autoencoders. Nat. Biotechnol. *39*, 555–560. https://doi.org/10.1038/s41587-020-00777-4.

143. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605–607. https://doi.org/10.1093/bioinformatics/btv638.

144. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. *3*, 836–843. https://doi.org/10.1038/s41564-018-0171-1.

145. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). MetaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834. https://doi.org/10.1101/gr.213959.116.

146. Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat. Biotechnol. *35*, 316–319. https://doi.org/10.1038/nbt.3820.

147. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. *48*, D265–D268. https://doi.org/10.1093/nar/gkz991.

148. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.

149. Pruitt, K.D.D., Tatusova, T., and Maglott, D.R.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, D61–D65. https://doi.org/10.1093/nar/gkl842.

150. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). GenBank. Nucleic Acids Res. *41*, D36–D42. https://doi.org/10.1093/nar/gks1195.

151. Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

152. Holm, L., and Laakso, L.M. (2016). Dali server update. Nucleic Acids Res. *44*, W351–W355. https://doi.org/10.1093/NAR/GKW357.

153. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. *36*, W5–W9. https://doi.org/10.1093/nar/gkn201.

154. Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248–254. https://doi.org/10.1016/0003-2697(76)90527-3.

155. Fish, W.W. (1988). Rapid colorimetric micromethod for the quantitation of complexed iron in biological samples. In Methods in Enzymology (Academic Press), pp. 357–364. https://doi.org/10.1016/0076-6879(88)58067-9.

156. Lorenzi, M., Gamache, M.T., Redman, H.J., Land, H., Senger, M., and Berggren, G. (2022). Light-driven [FeFe] hydrogenase based H_2 production in *E. coli*: a model reaction for exploring *E. coli* based semiartificial photosynthetic systems. ACS Sustain. Chem. Eng. *10*, 10760–10767. https://doi.org/10.1021/acssuschemeng.2c03657.

157. Senger, M., Eichmann, V., Laun, K., Duan, J., Wittkamp, F., Knör, G., Apfel, U.-P., Happe, T., Winkler, M., Heberle, J., et al. (2019). How [FeFe]-hydrogenase facilitates bidirectional proton transfer. J. Am. Chem. Soc. *141*, 17394–17403. https://doi.org/10.1021/jacs.9b09225.

158. Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams, K.H., Tringe, S.G., and Banfield, J.F. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome *1*, 22. https://doi.org/10.1186/2049-2618-1-22.

159. Minh, B.Q., Nguyen, M.A.T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. *30*, 1188–1195. https://doi.org/10.1093/molbev/mst024.

160. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli BL21(DE3) competent cells	Merck	69450-4
E. coli Origami B(DE3)™ competent cells	Merck	70837-3
'*Ca.* Lokiarchaeum ossiferum' B35 enrichments	This study	N/A
Chemicals, peptides, and recombinant proteins		
LB Broth	Sigma	L3022-1KG
Ampicillin	Sigma	A0166-5G
Di-sodium hydrogen phosphate	Sigma	1.06575.1000
Potassium dihydrogen phosphate	VWR	1.04873.1000
Sodium chloride	Sigma	S7653-1KG
Ammonium chloride	Sigma	213330-500G
Magnesium sulfate	VWR	M006
Calcium chloride	Merck	C5080-1KG
Glucose	Sigma	G8270-5KG
IPTG	VWR	437145X
Iron(II) sulfate	Scharlau	HI03510500
Tris	VWR	0497-1KG
Triton X-100	Sigma	T9284-100ML
Lysozyme	Sigma	62971–10g
DNase	Sigma	DN25-100mg
RNase	Sigma	R6513-50mg
$(Et_4N)_2[Fe_2(SCH_2NHCH_2S)(CO)_4(CN)_2]$ ($[2Fe]^{adt}$)	Synthesized in-house (Lab PI: Gustav Berggren) in accordance to literature protocols with minor modifications and verified by FTIR spectroscopy[95–99]	N/A
Di-potassium hydrogen phosphate	VWR	26932.290
Methyl viologen	Sigma	856177-1g
Sodium dithionite	Sigma	71699-250G
Kanamycin	VWR	TCIAK0047-5G
Magnesium chloride	Merck	1725711000
cOmplete™ EDTA-free protease inhibitor cocktail	Sigma	5892791001
StrepTrap XT	Cytiva	29401323
Biotin	Sigma	B4501-10G
Amicon 30 kDa Small Centricons	Sigma	UFC503096
Superdex 200 Increase 10/300 GL	Sigma	GE28-9909-44
Coomassie Plus (Bradford) Assay Reagent	ThermoFisher	10495315
Bovine Serum Albumin	Sigma	P0834-10X1ML
Fe^{2+} standard for AAS TraceCERT	Sigma	16596-250ML
Dithiothreitol	VWR	M109-5G
Ammonium iron(II) sulfate	Sigma	215406-100G
L-Cysteine	Sigma	C7352-100G
Recombinant cysteine desulfurase E. coli IscS	Kindly provided by Prof. Marc Fontecave and Dr. Sandrine Ollagnier de Choudens, CEA Grenoble[100]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PD-10 desalting column	VWR	17-0851-01
MES	Fisher	15432958
CHES	Sigma	C2885-25G
HEPES	Sigma	H3375-500G
TAPS	Sigma	T5130-100G
Sodium acetate	Sigma	S2889-250G
Polymyxin B Sulfate	Sigma	5291-1GM
Eosin Y	Sigma	230251-25G
TEOA	Sigma	90279-100ML
Sodium sulfate	Sigma	239313-500G
P1200 Grit Very Fine Sanding Sheet, 280mm x 230mm	RS PRO	188-3390
NucleoSpin Soil DNA extraction kit	Macherey-Nagel	740780.250
mirVana miRNA isolation kit	Invitrogen	AM1561
TURBO DNase	Invitrogen	AM2239
iST lysis buffer	Preomics	NC1699276
Deposited data		
Genome Taxonomy Database	Parks et al.[101]	GTDB 06-RS202 (https://gtdb.ecogenomic.org/)
PATRIC	Wattam et al.[102]	https://www.bv-brc.org/
HydDB	Søndergaard et al.[18]	https://services.birc.au.dk/hyddb/
40 assembled [FeFe] hydrogenase encoding archaeal MAGs	This paper	Figshare (dx.doi.org/10.6084/m9.figshare.25587258)
Proteomes of '*Ca.* Lokiarchaeum ossiferum' B35 enrichments	This study	PRIDE Archive (PXD047696)
Transcriptomes of '*Ca.* Lokiarchaeum ossiferum' B35 enrichments	This study	NBCI (BioProject PRJNA1054498)
Archaeal hydrogenase classification, structural modelling, and ATR-FTIR data and description	This study	Figshare (dx.doi.org/10.26180/25590891)
Recombinant DNA		
Genes encoding Mu, Ia, Ps, Na, Fm were codon optimized for expression in *E. coli*, and cloned in pET-11a(+) using restriction sites NdeI and BamHI	Genscript Gene Synthesis Services. See codon-optimized sequences in Table S5	N/A
Software and algorithms		
AlphaFold multimer v2.1.1	Tunyasuvunakool et al.[103]	N/A
BLAST+ v2.9.0	Camacho et al.[104]	N/A
Bowtie v2.5.1	Langmed and Salzberg[105]	N/A
CD-HIT v4.6	Fu et al.[106]	N/A
CheckM	Parks et al.[107]	N/A
Clustal Omega v1.2.2	Sievers and Higgins[108]	N/A
DIAMOND v0.9.31	Buchfink et al.[109]	N/A
dRep v.3.3	Olm et al.[110]	N/A
ETE v3.0.0	Huerta-Cepas et al.[111]	N/A
gggenes v0.4.1	https://github.com/wilkox/gggenes	N/A
GOOSOS	https://github.com/jwestrob/GOOSOS	N/A
GTDB-Tk v1.4.0	Chaumeil et al.[112]	N/A
HMMER v3.2.1	Wheeler and Eddy[113]	N/A
IQ-TREE v1.6.12	Nguyen et al.[114]	N/A
iTOL v6.3.2	Letunic and Bork[115]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
KofamKOALA v1.3.0	Aramaki et al.[116]	N/A
MAFFT v7.304	Katoh and Standley[117]	N/A
METABOLIC	Zhou et al.[118]	N/A
MMseqs2 v13-45111	Steinegger and Söding[119]	N/A
MsReport v0.0.14	Hollenstein et al.[120]	N/A
Pfamscan	Mistry et al.[121]	N/A
Prodigal v2.6.3	Hyatt et al.[122]	N/A
PSORTb v3.0.3	Yu et al.[123]	N/A
QT-PISA	Krissinel[124]	N/A
Spectronaut v17.6	https://biognosys.com/software/spectronaut/	N/A
StringTie v2.2.1	Perea et al.[125]	N/A
trimAl	Capella-Gutiérrez et al.[126]	N/A
Trimmomatic v0.36	Bolger et al.[127]	N/A
Geneious Basic	Kearse et al.[128]	https://www.geneious.com; RRID:SCR_010519
GPES software v4.9.006	Metrohm/Autolab	N/A
OPUS software v8.2 Build:8,2,21 (20181203)	Bruker Optik GmbH[129]	
Origin 8	Northampton, MA 01060 US	https://www.originlab.com/; RRID:SCR_014212
PGSTAT10	Eco/Chemie	N/A
Other		
ATR unit was sealed with a custom build PEEK cell that allowed for gas exchange and illumination mounted in a FTIR spectrometer	BioRadII from Harrick; Vertex V70v, Bruker[129]	N/A
Cell Kit for Rotating Electrodes (1x 24/25, 4x 14/20), 150 mL, water-jacketed	Equilabrium SAS	AKCELL3
Bearing Assembly (fits AFE6MB or AFE9MBA shafts to 24/25 joint)	Equilabrium SAS	AC01TPA6M
MSR Shaft for E5/E5TQ/E5HT RDE Tips and E6/E7/E8 RRDE Tips	Equilabrium SAS	AFE6MB
E5 RDE Tip, 5mm OD edge plane PG (epoxy encapsulated), 15mm OD PEEK shroud	Equilabrium SAS	AFE5T050GEPK
Graphite Counter Electrode Kit, 14/20 sleeve	Equilabrium SAS	AFCTR3B
Glass fritted tube, 12 mm OD, 14/20 Port	Equilabrium SAS	RRPG097
PTFE Adapter, 12 mm OD tube to 14/20	Equilabrium SAS	ACEP1420R12
Reference Electrode (silver chloride, single jxn, pin connector, 14/20 adapter)	Equilabrium SAS	RREF0021

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chris Greening (chris.greening@monash.edu).

Materials availability

Plasmids and bacterial strains generated in this study are available upon request from the Berggren Group, Uppsala Universitet.

Data and code availability

- Archaeal hydrogenase classification, structural modelling, and ATR-FTIR data and description have been deposited to Figshare (https://doi.org/10.26180/25590891). Assembled genomes reported can be accessed at Figshare (https://doi.org/10.6084/m9.figshare.25587258). Mass spectrometric proteomics and transcriptomics data of Ca. Lokiarchaeum ossiferum B35

enrichment culture have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD047696 and NCBI under BioProject ID PRJNA1054498, respectively.
- There is no original code reported in this publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Strains, media and growth conditions

'*Ca.* Lokiarchaeum ossiferum' B35 enrichments for transcriptomic and proteomic analyses were grown in MLM medium[53] at 20 °C under a headspace atmosphere of 80:20 N_2:CO_2 (0.3 bar). Per litre, the medium contains 20.7 g NaCl, 5 g $MgCl_2 \cdot 6H_2O$, 2.7 g $NaHCO_3$, 1.36 g $CaCl_2 \cdot 2H_2O$, 0.54 g NH_4Cl, 0.14 g KH_2PO_4, 0.03 g $Na_2S \cdot 9H_2O$, 0.03 g cysteine·HCl, 0.5 ml of acid trace element solution, 0.5 ml of alkaline trace element solution, 1 ml Se/W solution, and 0.1% casein hydrolysate (w/v). The acid trace element solution (per litre) contains 1.491 g $FeCl_2 \cdot 4H_2O$, 0.062 g H_3BO_3, 0.068 g $ZnCl_2$, 0.017 g $CuCl_2 \cdot H_2O$, 0.099 g $MnCl_2 \cdot 4H_2O$, 0.119 g $CoCl_2 \cdot 6H_2O$, 0.024 g $NiCl_2 \cdot 6H_2O$, and 4.106 ml HCl (37%). The alkaline trace element solution (per litre) contains 0.017 g Na_2SeO_3, 0.033 g Na_2WO_4, 0.021 g Na_2MoO_4, and 0.4 g NaOH. The medium pH was adjusted to 7.5 and contained ampicillin, kanamycin, and streptomycin (200 μg ml^{-1} each). The cultures were sampled every 14 days and DNA was extracted using the NucleoSpin Soil DNA extraction kit (Macherey-Nagel) according to the manufacturer's instructions. Growth was monitored by qPCR assays using primers LkF (5′- ATC GAT AGG GGC CGT GAG AG) and LkR (5′- CCC GAC CAC TTG AAG AGC TG) targeting lokiarchaeal 16S rRNA genes as previously described.[53]

Chemically competent *E. coli* BL21(DE3) and *E. coli* Origami™ B(DE3) strains were used as vectors for heterologous expression of archaeal [FeFe] hydrogenases. Starter cultures were grown overnight in LB medium containing 100 μg mL^{-1} ampicillin at 37°C. These cultures were subsequently used to inoculate M9 medium (22 mM Na_2HPO_4, 22 mM KH_2PO_4, 85 mM NaCl, 18 mM NH_4Cl, 0.2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (v/v) glucose) containing 100 μg/mL ampicillin and 15 μg/mL kanamycin for protein expression and purification. Cultures were grown at 37°C and 150 rpm.

METHOD DETAILS

Genome sources

In this study, 130 archaeal genomes encoding [FeFe] hydrogenases were analysed. 40 genomes are MAGs retrieved from our unpublished datasets at ggKbase (https://ggkbase.berkeley.edu/archaeal_hydrogenases). 77 genomes were retrieved from the Genome Taxonomy Database (GTDB) R06-RS202[101] following a search of all 2,339 archaeal species representative genomes. GTDB was chosen as the data source as it offers a standardised taxonomy, a manageable search space due to pre-clustered sequences, and genomes pre-annotated with gene predictions.[130] In addition, 13 other Asgardarchaeota genomes were retrieved from PATRIC.[102] Full details of the genomes analysed are provided in Table S1. All 130 genomes were examined for completeness and contamination with CheckM[107] and manually curated through taxonomic profiling in ggKbase. Initial taxonomic classification of genome bins was performed with GTDB-Tk-R202[112] and subsequently confirmed by constructing phylogenetic trees. All archaeal genomes analysed were dereplicated at 95% average nucleotide identity using dRep v.3.3.[110]

Metagenomic assembly and binning

The 40 MAGs were retrieved from a range of anoxic ecosystems. Most MAGs came from previously reported study sites, namely Guaymas Basin hydrothermal vents (7 MAGs; Gulf of California, Mexico),[131] Lac Lavin freshwater lake (6 MAGs; central France),[132] Crystal Geyser (6 MAGs; Utah, USA),[133] Chinese hot springs (5 MAGs; 2 from Yunnan Province and 3 from Tibetan Plateau, China),[134] an aquifer (3 MAGs; Napa County, California, USA),[135] Manure lagoon (2 MAGs, California, USA),[135] an aquifer adjacent to the Colorado River (3 MAGs; Rifle, Colorado, USA),[136] a wetland soil (2 MAGs; Napa County, California, USA),[137] Alum Rock mineral spring (1 MAG; San Jose, California),[138] Zodletone Spring (1 MAG; Anadarko, Oklahoma, USA), Azore Islands hot springs (1 MAGs, Portugal),[139] a borehole (1 MAG, Muzunami, Japan).[139] Sampling, DNA extraction, sequencing library preparation, and sequencing methods were previously described. Briefly, metagenomic sequencing reads assembled using IDBA-UD.[140] Contigs larger than 2.5 kb were retained, and sequencing reads from all samples were mapped against each resulting assembly utilizing Bowtie2.[105] Differential coverage profiles, filtered with a 95% read identity threshold, were then used for genome binning using a suite of binning tools (MetaBAT2,[141] VAMB,[142] MaxBin2,[143] Abawaca (https://github.com/CK7/abawaca), with the final bin choice determined by DAS Tool.[144] Additionally, 1 MAG was obtained from samples from Corona Mine drainage at the Oat Hill Mine, Napa County, California, USA; in this case, water was filtered through 2.5 and 0.1 μm filters sequentially, DNA was extracted from each filter separately, and two runs of Illumina paired-end sequencing were performed at 150 bp and 250 bp read lengths. 1 MAG were also obtained from hyporheic zone water sampled from beneath the riverbed of the East River, Gunnison County, Colorado, USA; in this case, water was filtered through a 0.1 μm filter, DNA was extracted from the filter, and Illumina sequencing was performed with a read length of 150 bp. For the East River and Oat Hill Mine samples, genome data were assembled using MetaSPAdes v3.15.5[145] with default k-mer values. Draft genomes reconstructed using the MetaBAT2,[141] VAMB,[142] MaxBin2,[143] and via ggKbase manual binning tools and the best genome selected using DAS Tool.[144]

Identification of archaeal [FeFe] hydrogenase genes

The process of identifying [FeFe] hydrogenase enzymes in archaea involved matching a "training" profile of known [FeFe] hydrogenases against a database of "candidate" protein sequences. Representative archaeal genomes from GTDB R06-RS202 were used as the data source for candidate hydrogenases.[101] The training data was the catalytic domains of the [FeFe] hydrogenases identified in the HydDB.[18] The analysis pipeline was written using the Nextflow pipeline framework, which allowed for the analysis to be run reproducibly in containers, which were executed in parallel across nodes of the computing cluster.[146] The first process in the pipeline performed a multiple sequence alignment of training sequences, which was then used to build a hidden Markov model (HMM) using hmmer3[113] that modelled the primary sequence of the [FeFe] hydrogenase catalytic domain. Candidate protein sequences with length greater than 100,000 amino acids were first excluded from the analysis. Each protein sequence from each representative genome was then matched against the HMM to obtain a bit score, which represented the similarity of the candidate sequence to the known enzyme profile. Preliminary matches with a positive bit score were retained and then filtered by the presence of the CxxxC motif required to ligate the [FeFe] hydrogenase catalytic centre. The matches passing this filter were then aggregated into a report, along with their taxonomy and bit score. Following this, the matches were manually inspected through a combination of Conserved Domain Database (CDD) annotations[147] and phylogenetic analysis to derive a bit score cutoff. A final cutoff of 15.9 was chosen to include all true positives and exclude all other sequences.

Analysis of [FeFe] hydrogenase domain architecture

[FeFe] hydrogenases are often multidomain proteins comprising a conserved H_2-activating domain named the H-cluster and various accessory domains at the N- and C-terminals involved in electron transfer or H_2 sensing.[1,37] To locate the positions of the H-cluster, all retrieved archaeal [FeFe] hydrogenase sequences were aligned against the trimmed reference [FeFe] hydrogenase H-cluster sequences from HydDB[18] using Clustal Omega v1.2.2 (default setting).[108] N- and C-terminal regions outside of the H-cluster were extracted and annotated against CDD v3.19[147] using rpsblast (-evalue 0.01 -max_hsps 1 -max_target_seqs 10) in BLAST+ v2.9.0.[104] The N-terminal fusion of the small subunit of [NiFe] hydrogenase with certain archaeal [FeFe] hydrogenases was confirmed by searching against Pfam protein family database v34.0[148] and protein structural modelling as detailed below. The N- and C-terminal sequences of [FeFe] hydrogenases were further searched for the presence of signature iron-sulfur cluster motifs or specific pattern of cysteine residues that are potentially involved in electron transfer: plant ferredoxin-like [2Fe2S] ($C_{x10-11}C_{x2}C_{x11-15}C$); bacterial ferredoxin-like 2[4Fe4S] ($C_{x2}C_{x2}C_{x3}C_{x23-32}C_{x2}C_{x2}C_{x3}C$); $(Cys)_3$His-ligated [4Fe4S] ($H_{x2-3}C_{x2}C_{x5}C$).[70] Based on the domain organization and the classification scheme by Land et al.[37], archaeal [FeFe] hydrogenases were assigned into different subclasses. Domain annotation and subclass classification results of the [FeFe] hydrogenases are summarized in Table S2.

Analysis of [FeFe] hydrogenase genetic organization

To characterise the genetic context and potential interacting proteins of archaeal [FeFe] hydrogenases, up to 10 genes upstream and downstream of the catalytic subunits were retrieved. These neighbouring genes were annotated against CDD v3.19[147] using rpsblast (-evalue 0.01 -max_hsps 1 -max_target_seqs 5) in BLAST+ v2.9.0,[104] the Pfam protein family database v34.0[148] using PfamScan v1.6 (default setting),[121] and the NCBI RefSeq protein database release 202[149] using DIAMOND v0.9.31 blastp algorithm (–max-hsps 1 –max-target-seqs 1).[109] Protein subcellular localization and the presence of internal helices of the gene were predicted using PSORTb v3.0.3 (–archaea).[123] The subgroup lineage of the flanking gene that encodes large subunit of [NiFe] hydrogenase was classified using HydDB.[18] To facilitate curation of the annotation results and identification of conserved neighbouring genes, all flanking genes were clustered at an identity threshold of 30% and a minimum coverage of 80% using MMseqs2 (–min-seq-id 0.3 -c 0.8 –cov-mode 1 –cluster-mode 2 -s 7.5).[119] The R package gggenes v0.4.1 (https://github.com/wilkox/gggenes) was used to construct gene arrangement diagrams. All sequences, annotations, and genetic arrangements are summarized in Table S2.

Identification of archaeal [FeFe] hydrogenase maturases

To probe the maturation pathway and evolution of [FeFe] hydrogenases in Archaea, we performed a genomic survey of the three conserved [FeFe] hydrogenase maturases (HydE, HydF, HydG) in all representative archaeal species in GTDB R06-RS202.[101] We first compiled a comprehensive database of known [FeFe] hydrogenase maturase sequences from bacteria and eukaryotes and expanded the dataset based on BLAST searches against the NCBI non-redundant protein database (Nov 2021).[150] New and authentic divergent hits including from archaea were included in the final database, which was then used to screen for the presence of [FeFe] hydrogenase maturases in archaeal genomes. To enable the discovery of phylogenetically novel maturases, a relaxed default setting of the DIAMOND v0.9.31 BLASTp algorithm was first applied.[109] False positive hits were filtered by further searching against CDD v3.19[147] for the presence of rSAM_HydE (TIGR03956), rSAM_HydG (TIGR03955), and GTP_HydF (TIGR03918) domains. The [FeFe] hydrogenase maturase databases and hits from the archaeal genomes are provided in Table S3.

'*Ca.* Lokiarchaeum ossiferum' transcriptomic analysis

When '*Ca.* L. ossiferum' B35 cells reached mid to late exponential growth (~8.0×10^6 16S rRNA gene copies/mL), 35 ml of enrichment culture was centrifuged at 20,000 × *g* for 30 minutes, 4°C. The resulting pellet was used for RNA extraction, which was performed using the mirVana miRNA isolation kit (Invitrogen), according to the manufacturer's instructions. To remove any leftover genomic DNA, the samples were incubated with TURBO DNase (Invitrogen) at 37°C for 1 h. The absence of DNA was confirmed

through lokiarchaeal 16S rRNA gene PCR assays. Metatranscriptomic analysis was performed using the NovaSeq PE150 platform through the services of the company Novogene (United Kingdom). The RNA-seq data were analyzed with Trimmomatic v0.36[127] to remove adaptor sequences and filtering low-quality reads using the following parameters: LEADING:3 TRAILING:3 MINLEN:36. High quality reads from both replicates (s1 and s2) were mapped to the genome of '*Ca.* L. ossiferum' B35 using bowtie2 v2.5.1.[105] Transcript expression was normalized into transcripts per million (TPM) using StringTie v2.2.1.[125]

'*Ca.* Lokiarchaeum ossiferum' proteomic analysis

Pellets of exponential growth phase '*Ca.* L. ossiferum' enrichments were suspended in 100 μL iST lysis buffer (Preomics). To lyse cells, the suspensions were heated (95°C, 1,000 rpm, 10 min), sonicated (water bath, 3 mins), homogenized (Beatbox (Preomics), 10 mins, high setting), and further heated (95°C, 1,000 rpm, 10 min). Samples were then centrifuged at 10,000 × *g* for 1 min and 50 μL supernatants were digested using the iST kit following the vendor's protocol (Preomics). Peptides were separated on a Vanquish Neo nano-flow chromatography system (Thermo-Fisher), using a pre-column for sample loading (Acclaim PepMap C18, 2 cm × 0.1 mm, 5 μm, Thermo Fisher), and a C18 analytical column (Acclaim PepMap C18, 50 cm × 0.75 mm, 2 μm, Thermo-Fisher), applying a segmented linear gradient from 2% to 35% and finally 80% solvent B (80 % acetonitrile, 0.1 % formic acid; solvent A 0.1 % formic acid) at a flow rate of 230 nL min^{-1} over 120 min. Eluting peptides were analyzed on an Exploris 480 Orbitrap mass spectrometer (Thermo Fisher), which was coupled to the column with a FAIMS pro ion-source (Thermo-Fisher) using coated emitter tips (PepSep, MSWil). The mass spectrometer was operated in DIA mode with the FAIMS CV set to -45, the survey scans were obtained in a mass range of 350-1200 m/z, at a resolution of 60k at 200 m/z and a normalized AGC target at 300%. 60 MSMS spectra with variable isolation width between 7 and 191 m/z covering 349.5-1200.5 m/z range, including 1 m/z windows overlap, were acquired in the HCD cell at 30% collision energy at a normalized AGC target of 1000% and a resolution of 30k. The max. injection time was set to auto. Raw data were processed using Spectronaut software v.17.6 (https://biognosys.com/software/spectronaut/) with the DirectDIA+ workflow. A custom database that included the *Ca.* L. ossiferum B35 proteins and the most common contaminants was used. The searches were performed with full trypsin specificity and a maximum of 2 missed cleavages at a protein and peptide spectrum match false discovery rate of 1%. Carbamidomethylation of cysteine residues were set as fixed, oxidation of methionine and N-terminal acetylation as variable modifications. The cross-run normalization was turned off and all other settings were left as default. Computational analysis was performed using Python and the in-house developed Python library MsReport v.0.0.14.[120] Protein intensity less than 1000 was set to missing to remove the low-quality quantification values. LFQ protein intensities reported by SpectroNaut were log_2-transformed and normalized across samples using the ModeNormalizer from MsReport. The missing normalized LFQ intensity values were imputed by drawing random values from a normal distribution after filtering out contaminants, proteins with less than 2 peptides and less than 1 quantified values in at least one group.

AlphaFold2 structural modelling

[FeFe] hydrogenase models were generated using AlphaFold multimer v2.1.1 implemented on the Monash University MASSIVE M3 computing cluster.[65,103] The amino acid sequences for HydA from each of the [FeFe] hydrogenase groups (A1, A3, B, E, and F) shown in Table S5 and below were modelled both alone and with the sequences of putative complex partners present in the hydrogenase gene cluster. Modelling with putative complex partners was performed iteratively, with output models assessed to determine if a credible complex was generated. Predicted complexes were then modelled with higher stoichiometries, where possible given limitations of GPU RAM (~2,500 amino acids), to predict larger order structures. Models produced were validated based on confidence scores (pLDDT) with only regions with a confidence score of >85 utilised for analysis. Where complexes were predicted subunit interfaces were inspected manually for surface complementarity and the absence of clashing atoms. Interfaces were also analysed for stability using the program QT-PISA,[124,151] with only interfaces predicted to be stable utilised for analysis. To assign cofactors to [FeFe] hydrogenase models generated by AlphaFold the closest homologous structures or domains were identified by searching the PDB database using NCBI BLAST or the DALI server.[152,153] The homologous structures were aligned with the AlphaFold models and cofactors were added in corresponding positions to that of the experimental structures, providing all conserved coordinating residues were present. Cofactor position was then manually adjusted to optimise coordination and to minimise clashes. The PDB IDs for the structures were used to model cofactors for the archaeal hydrogenases are:

- Group A1 – UBA95 sp002499405 (Micrarchaeota), '*Ca.* Iainarchaeum andersonii': HydA = 6GM0 Chain A
- Group A3 – DSAL01 sp011380095 (Altarchaeota): HydA = 6GM0 Chain A; HydB (NuoF-like) = 7E5Z chain B; HydC (NuoE-like) = 7E5Z chain A
- Group B – '*Ca.* Prometheoarchaeum syntrophicum': HydA = 6GM0 Chain A ([FeFe] hydrogenase domain), 7BKD Chain A (ferredoxin domain); HydC (NuoE-like) = 8E9G Chain E
- Group E – CABMGN01 sp902385635 (Nanoarchaeota), '*Ca.* Forterrea multitransposorum': HydA = 6GM0 Chain A
- Group F – UBA147 sp002496385 (Thermoplasmatota): HydA = 6GM0 chain A ([FeFe] hydrogenase domain), 5ODC chain E ([NiFe] hydrogenase small domain); HyhL = 5ODC chain L; HydD (NuoG-like) = 7T2R chain A, 5ODC chain E; HydB (NuoF-like) = 7E5Z chain B, 7MFM chain I; HydC (NuoE-like) = 7E5Z chain A

Protein expression and characterization

All chemicals used during the protein production and characterization were purchased from VWR and used as received unless otherwise stated. The genes encoding group A, B, E, and F [FeFe] hydrogenases (Table S1) were codon optimized for expression in *E. coli*, synthesized, and cloned in pET-11a(+) by Genscript using restriction sites NdeI and BamHI. Codon optimized sequences can be found in Table S5. Expression constructs were re-transformed in chemically competent *E. coli* BL21(DE3) cells to express the apo-forms of the hydrogenases lacking the diiron subsite of the H-cluster. Starter cultures were grown overnight in 5 mL LB medium containing 100 μg mL^{-1} ampicillin at 37°C. These cultures were subsequently used to inoculate 80 mL of M9 medium (22 mM Na_2HPO_4, 22 mM KH_2PO_4, 85 mM NaCl, 18 mM NH_4Cl, 0.2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (v/v) glucose) containing 100 μg mL^{-1} ampicillin. Cultures were grown at 37°C and 150 rpm until an optical density (OD_{600}) of approximately 0.4 to 0.6 was reached. Protein expression was induced by the addition of 0.1 mM $FeSO_4$ and 1 mM IPTG. Induced cultures were incubated at 20°C and 150 rpm for approximately 16 h. Cells were thereafter harvested by centrifugation at 4,930 × *g* for 10 mins at 4°C. All subsequent operations were carried out under anaerobic conditions to prevent hydrogenase inactivation by atmospheric oxygen in an MBRAUN glovebox ($[O_2] < 5$ ppm). The cell pellet was resuspended in a 0.5 mL lysis buffer (30 mM Tris-HC pH 8.0, 0.2 % (v/v) Triton X-100, 0.6 mg mL^{-1} lysozyme, 0.1 mg mL^{-1} DNase, 0.1 mg mL^{-1} RNase). Cell lysis was performed by three cycles of freezing/thawing in liquid N_2, and the supernatant was recovered by centrifugation (29,080 × *g*, 10 mins, 4°C).

Hydrogenase semisynthetic maturation

The subsite $[2Fe]_H$ subsite mimic, $(Et_4N)_2[Fe_2(SCH_2NHCH_2S)(CO)_4(CN)_2]$ ($[2Fe]^{adt}$), was synthesized in accordance to literature protocols with minor modifications and verified by FTIR spectroscopy.[95–99] Incorporation of cofactor was performed by addition of 100 μg the $[2Fe]^{adt}$ subsite mimic (final concentration 80 μM) to 380 μL of the supernatant in potassium phosphate buffer (100 mM, pH 6.8) and 1 % (v/v) Triton X-100. The reaction mixture was enclosed in an airtight vial and was anaerobically incubated at 20°C for 1-4 hr. The non-purified lysate containing the $[2Fe]^{adt}$ subsite mimic was mixed with 200 μL of potassium phosphate buffer (100 mM, pH 6.8) with 10 mM methyl viologen and 20 mM sodium dithionite. Reactions were incubated at 37°C up to 120 mins. H_2 production was determined by analysing the reaction headspace every 15 mins using a PerkinElmer Clarus 500 gas chromatograph (GC) equipped with a thermal conductivity detector (TCD) and a stainless-steel column packed with Molecular Sieve (60/80 mesh). The operational temperatures of the injection port, oven, and detector were 100°C, 80°C, and 100°C, respectively. Argon was used as carrier gas at a flow rate of 35 mL min^{-1}. The three biological replicates were run at varying times (1-4 hours) of incubating the cell lysates with the $[2Fe]^{adt}$ subsite mimic. Incubation time was not found to influence the observed H_2 production. Thus, variation in H-cluster formation rates did not appear to have a substantial influence on the outcome of the screening process.

Hydrogenase purification

Expression construct with verified sequence of the group E [FeFe] hydrogenase from '*Ca.* Forterrea multitransposorum' (Fm) was retransformed in chemically competent *E. coli* Origami™ B(DE3) cells since the strain yielded the highest expression levels from the small-scale expression tests. Starting cultures were grown overnight in 10 mL LB medium containing 100 μg/mL ampicillin and 15 μg/mL kanamycin at 37°C. These cultures were subsequently used to inoculate 1 L of M9 medium (22 mM Na_2HPO_4, 22 mM KH_2PO_4, 85 mM NaCl, 18 mM NH_4Cl, 0.2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (v/v) glucose) containing 100 μg/mL ampicillin and 15 μg/mL kanamycin. Cultures were grown at 37°C and 150 rpm until an optical density (OD_{600}) of approximately 0.4 was reached. Protein expression was induced by the addition of 0.1 mM $FeSO_4$ and 1 mM IPTG. Induced cultures were incubated at 20°C and 150 rpm for approximately 16 h. Cells were thereafter harvested by centrifugation in a Beckman Coulter Avanti J-25 centrifuge (4,424 × *g*, 10 min). All subsequent operations were carried out under anaerobic conditions in the glovebox to prevent hydrogenase inactivation by atmospheric oxygen. The cell pellet was resuspended in 100 mM Tris-HCl pH 8.0, with NaCl (150 mM), $MgCl_2$ (10 mM), lysozyme from chicken egg white (1 mg/mL), DNAse I from bovine pancreas (0.05 mg/mL), RNase A from bovine pancreas (0.05 mg/mL), and a tablet of cOmplete™ EDTA-free protease inhibitor cocktail, and was incubated inside the glovebox for 30 min. Cell lysis was performed by three cycles of freezing/thawing in liquid N_2. Cell debris was removed by centrifugation in a Beckman Coulter Optima L-90K Ultracentrifuge (222,592 × *g*, 60 min). The supernatant was collected and filtered (0.45 μm syringe filter) before being loaded on a StrepTrap™ XT (Cytiva) affinity column using a BioLogic DuoFlow™ FPLC system (Bio-Rad) and purified according to the manufacturerś instructions. Products eluted by 50 mM biotin were concentrated using Amicon®Ultra 30 kDa molecular weight cut-off (MWCO) centrifugal filters (Merck Millipore Ltd.). The StrepTrap™ eluates were further separated using size exclusion chromatography via Superdex™ 200 10/300 GL, equilibrated in 100 mM Tris-HCl, 150 mM NaCl pH 8.0, to acquire purified Fm (Figures S5B and S5C). Samples were stored anaerobically at −80 °C. Coomassie-stained SDS-PAGE was used to assess purity. Protein estimations were performed via Bradford assay using bovine serum albumin as a standard.[154] Quantification of Fe-content was performed using a previously reported assay[155] using a commercially available Fe^{2+} standard for AAS TraceCERT (Sigma Aldrich) for the calibration curve.

Activation of purified hydrogenase

To semi-enzymatically reconstitute the iron-sulfur clusters of Fm, a solution of 50 μM apoprotein in 100 mM Tris-HCl, 150 mM NaCl pH 8.0 was incubated with 500 μM dithiothreitol (DTT) under strictly anaerobic conditions for 10 min at room temperature. The iron and sulfur sources were ferrous ammonium sulfate and L-cysteine, respectively, both added in 1.5-fold molar excess to the desired

number of Fe-atoms to be added. Reconstitution was initiated by adding a 1% molar equivalent of recombinant cysteine desulfurase (*E. coli* IscS), slowly releasing sulfide *in situ* from cysteine. Reaction mixtures were incubated at room temperature up to 2 hours. At the same time, the increase of absorbance around 405 nm was monitored by UV/Vis (Figures S5B and S5C). The reconstitution process was stopped by running the reaction mixture through a PD-10 column (GE Healthcare), equilibrated in 100 mM Tris-HCl, 150 mM NaCl pH 8.0. To activate Fm with $[2Fe]^{adt}$, under strictly anaerobic conditions, the reconstituted enzyme (50 μM) was mixed with sodium dithionite (1 mM, 20× excess) in 100 mM phosphate buffer, pH 6.8 and incubated in room temperature for 10 minutes. Cofactor incorporation started with the addition of $[2Fe]^{adt}$ (600 μM, 12× excess), and the reaction mixture was incubated for 1 h. The mixture was loaded onto a PD-10 desalting column (GE Healthcare) equilibrated with 10 mM Tris-HCl pH 8.0. The sample was concentrated using Amicon®Ultra 30 kDa MWCO centrifugal filters (Merck Millipore Ltd.), aliquoted into PCR tubes, and transferred into airtight serum vials (3-5 μL each) before they were flash-frozen in liquid N_2 and stored at -80°C until further use.

Protein film electrochemistry

Protein film electrochemistry experiments were carried out under anaerobic conditions at 20°C and pH 7.0. The three-electrode system was made up of (1) Ag/AgCl (4 M KCl) as reference electrode, (2) rotating disk 5 mm OD pyrolytic graphite edge (PGE) plane (epoxy encapsulated) as working electrode, and (3) graphite rod as the counter electrode. The gas-tight glass cell used featured a water jacket for temperature control and a cell gas inlet/outlet for hydrogen flow control. The buffer used was composed of 5 mM MES, 5 mM CHES, 5 mM HEPES, 5 mM TAPS, 5 mM sodium acetate (NaOAc), with 0.1 M Na_2SO_4 as carrying electrolyte titrated with H_2SO_4 to pH 7.0, and purged with N_2 for 3 to 4 hours. The PGE working electrodes were polished with P1200 sandpaper and rinsed with purified water before they were brought into the glovebox. To remove residual O_2 in the PGE electrode, cyclic voltammograms were run at 100 mV/s from -100 to -600 mV (vs. standard hydrogen electrode (SHE)) for 40 scans. The cyclic voltammogram of the blank electrode (no enzyme immobilized) was then recorded at 10 mV/s with the working electrode rotated at 3 krpm. Polycationic polymyxin B sulfate (5 μL of 0.2 mg/mL) was added onto the deaerated PGE surface before adding 5 uL of 5 μM activated enzyme. The mixture was left for 10 min for maximal adsorption before the excess solution was removed by pipet. The cyclic voltammogram of the system with the immobilized enzyme was then recorded at 10 mV/s under 1 atm of H_2. Electrochemical data was acquired using an Eco/Chemie PGSTAT10 and the GPES software (Metrohm/Autolab). Data were analyzed using Origin 8 software. All values are referenced versus SHE. Experiments were conducted on two independent enzyme films, with each film scanned at least three times.

ATR-FTIR spectroscopy

2 μL enzyme solution (110 μM of Fm) in 10 mM Tris buffer (pH 8) was deposited on the ATR crystal. The ATR unit (BioRadII from Harrick) was sealed with a custom build PEEK cell that allowed for gas exchange and illumination mounted in a FTIR spectrometer (Vertex V70v, Bruker).[129] The sample was dried under 100% nitrogen gas and rehydrated with a humidified aerosol (100 mM Tris-HCl, pH 8). Spectra were recorded with 2 cm^{-1} resolution, a scanner velocity of 80 Hz and averaged of varying number of scans (mostly 1000 Scans). All measurements were performed at ambient conditions (room temperature and pressure, hydrated enzyme films). Photochemical reduction was achieved through a previously established protocol.[156,157] In short, an enzyme film prepared as described above also including Eosin Y (6 mM, 0.5 μL) and triethanolamine (TEOA, 200 mM, 2 μL) was illuminated using a Schott KL2500LCD cold light source. Spectra shown in Figure 3 are representative examples of two technical replicates. Data were analysed using OPUS[129] and Origin 2019 Software.

Genome-wide metabolic annotations

For all archaeal genomes encoding [FeFe] hydrogenases, genes were predicted using Prodigal v2.6.3.[122] Preliminary functional annotations were established and cross-referenced using KEGG HMMs, UniRef100 and UniProt, and collections of metabolic capacities in genome bins were reviewed using ggKbase genome summaries, as previously described.[56] Each open reading frame in the archaeal genomes was assigned a KEGG Orthology if the best-scoring KEGG HMM surpassed the bitscore cutoff. For a targeted profiling of the metabolic capacity of the archaeal genomes, we performed a homology-based search against our custom curated database (https://doi.org/10.26180/c.5230745) consisting of metabolic marker genes involved in carbon fixation (RbcL, AcsB, AclB, Mcr, Hbs), alternative electron donors (FdhA, CoxL, CooS, McrA, MmoA, PmoA, IsoA, FCC, Sqr, Sor, SoxB, PsaA, PsbA, ARO), alternative electron acceptors (AsrA, DsrA, NarG, NapA, NirS, NirK, NrfA, NosZ, NorB, Nod, MtrB, OmcB, YgfK, RdhA), respiration (SdhA, FrdA, CoxA, CcoN, CyoA, CydA, AtpA) using DIAMOND v.2.0.11 with filtering cutoffs described previously.[44] Briefly, hits were filtered based on a minimum query cover of 80% and an identity threshold of 50%, except for CoxL, MmoA, and RbcL (all 60%), AtpA, ARO, IsoA, PsbA, and YgfK (all 70%), Hbs (75%), and PsaA (80%). The functional annotations were expanded to include marker genes for fermentation, fatty acid degradation, aromatic compound degradation, carbohydrate metabolism, and sulfur metabolism using curated HMMs from METABOLIC[118] and KofamKOALA v1.3.0[116] with default bitscore thresholds. Hits were further inspected and filtered through searching the NCBI CDD database and final annotation results are summarized in Table S1.

Genome-level phylogenetic analysis

To construct the archaeal genome tree, 15 conserved syntenic ribosomal proteins[158] were retrieved, aligned, and concatenated using GOOSOS (https://github.com/jwestrob/GOOSOS) using all archaeal genomes at least 60% complete and less than 5%

contaminated (based on CheckM[107]). Genomes containing at least 75% of the 15 syntenic proteins were retained. The concatenated ribosomal protein sequences were aligned using MAFFT,[117] followed by trimming with trimAl[126] using the -gt 0.1 option. The final length of the trimmed concatenated protein alignment was 3224 amino acids for 118 genomes. Branch support was obtained using the ultrafast bootstrap method[159] implemented in IQ-TREE v1.6.12[114] and the phylogeny was estimated utilizing the following parameters -bb 1000 -m LG+F+G4. All trees were visualized using iTOL v6.3.2.[115]

Gene-level phylogenetic analysis

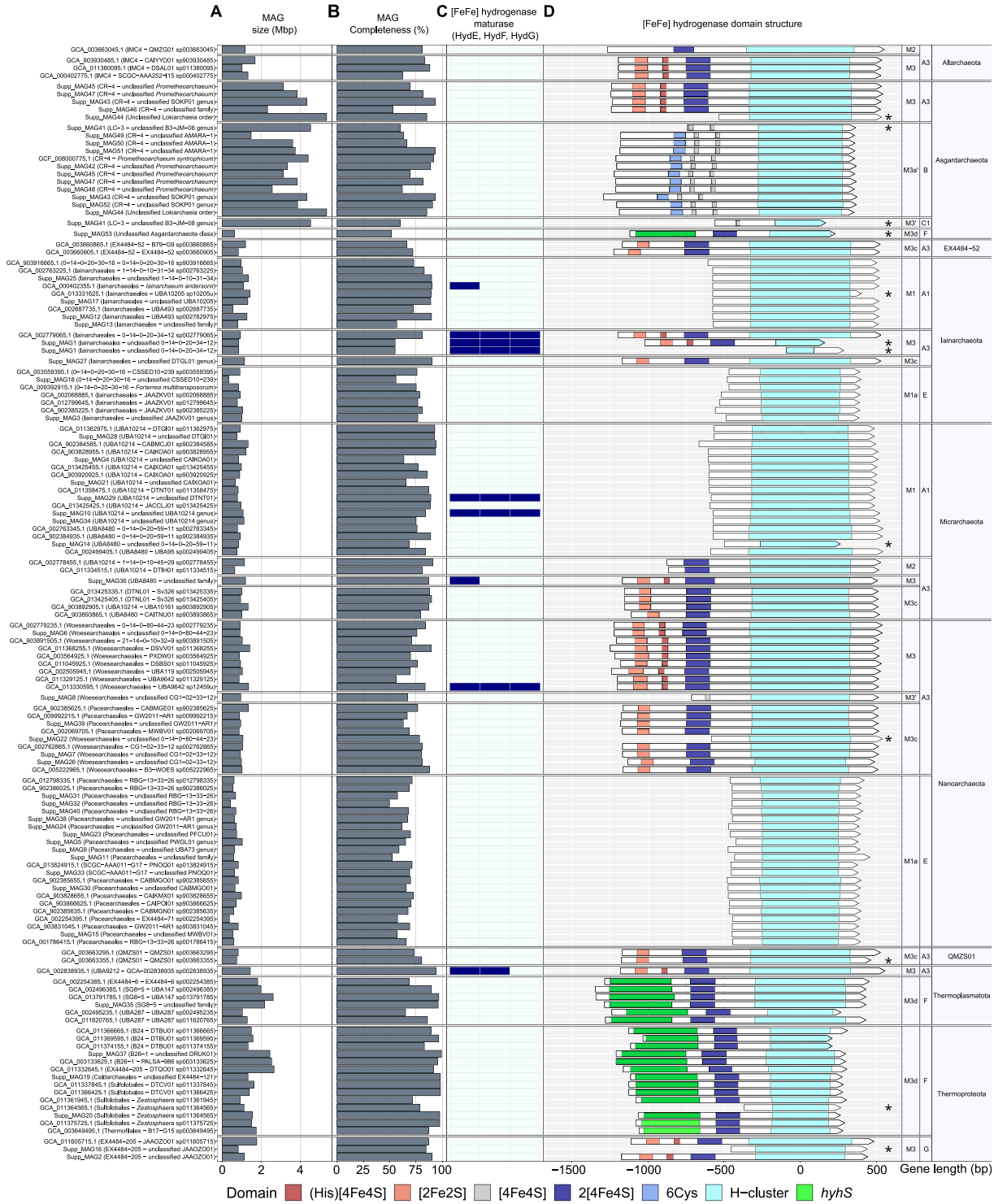
The amino acid sequences of [FeFe] hydrogenases catalytic subunits (HydA) were retrieved from three datasets: the archaeal genomes analysed in this study, all reference sequences from the hydrogenase database (HydDB),[18] and additional eukaryotic genomes.[89] For all datasets, taxonomy was assigned to each sequence using ETE v3.0.0[111] and CD-HIT v4.6[106] was used to reduce the dataset at the 80% amino acid sequence identity level. The multiple sequences retrieved were aligned using MAFFT v7.304 (settings: –localpair –maxiterate 1000 –reorder).[117] The resulting alignment was trimmed using trimAl (settings: -gt 0.1)[126] and manually inspected with Geneious[128] to remove partial sequences (Dataset S1). Using ModelFinder with default parameters implemented in IQ-TREE v1.6.1,[114] the best-fit model according to Bayesian information criterion (BIC) was determined to be LG+C20+R+F. Maximum likelihood phylogenetic trees using the LG+F+G4 substitution model were constructed using IQ-TREE v1.6.1[114] (setting: -m LG+C20+R+F -nt AUTO -bb 1000 -alrt 1000). To ensure the robustness of the phylogenetic inference, trees were built with 1000 ultrafast bootstraps and 1000 aLRT (Approximate Likelihood Ratio Test)[160] (Dataset S1). The models were applied to 3,677 amino acid sequences of the catalytic subunit (HydA) of [FeFe] hydrogenases, including the hybrid hydrogenases.

QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical comparison was made in this study. The experimental data were shown as the mean ± SEMs and the number of replicates was specified in figure legends where appropriate. No data were excluded for data analysis.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Genome statistics, hydrogenase maturases, and hydrogenase domain structure of 130 [FeFe] hydrogenase-encoding archaeal genomes, related to Figures 1, 2, 4, and 5

(A and B) Bar charts showing the size and completeness (CheckM) of [FeFe] hydrogenase-encoding archaeal genomes from GTDB R05-RS202 (77 in total), newly assembled metagenomes (40 in total), and PATRIC (13 in total).

(C) Heatmap showing the detection of known [FeFe] hydrogenase maturase (HydEFG) homologs based on homology search (Table S3). Navy shading denotes the presence of the homolog.

(D) Domain and iron-sulfur cluster organization of archaeal [FeFe] hydrogenase.

Incomplete open reading frames are denoted by asterisks (*) next to arrows. Subclass (protein domain structure-based scheme) and subgroup (protein phylogeny-based scheme) classification of [FeFe] hydrogenases are based on Land et al.[37] and Greening et al.,[1] respectively, with modifications proposed by the current study. H-cluster, catalytic domain of [FeFe] hydrogenase; *hyhS*, [NiFe] hydrogenase small subunit/iron-sulfur domain; (His)[4Fe4S], $(Cys)_3$His-ligated [4Fe4S] cluster binding domain; [2Fe2S], [2Fe2S] cluster binding domain; [4Fe4S], [4Fe4S] cluster binding domain; 2[4Fe4S], bacterial ferredoxin-like 2[4Fe4S] cluster binding domain; 6Cys, putative iron-sulfur cluster binding domain. Note that the phylum QMZS01 was classified as Aenigmatarchaeota in GTDB R06-RS207, while Thermoproteota class EX4484-205 was proposed as Brockarchaeia.

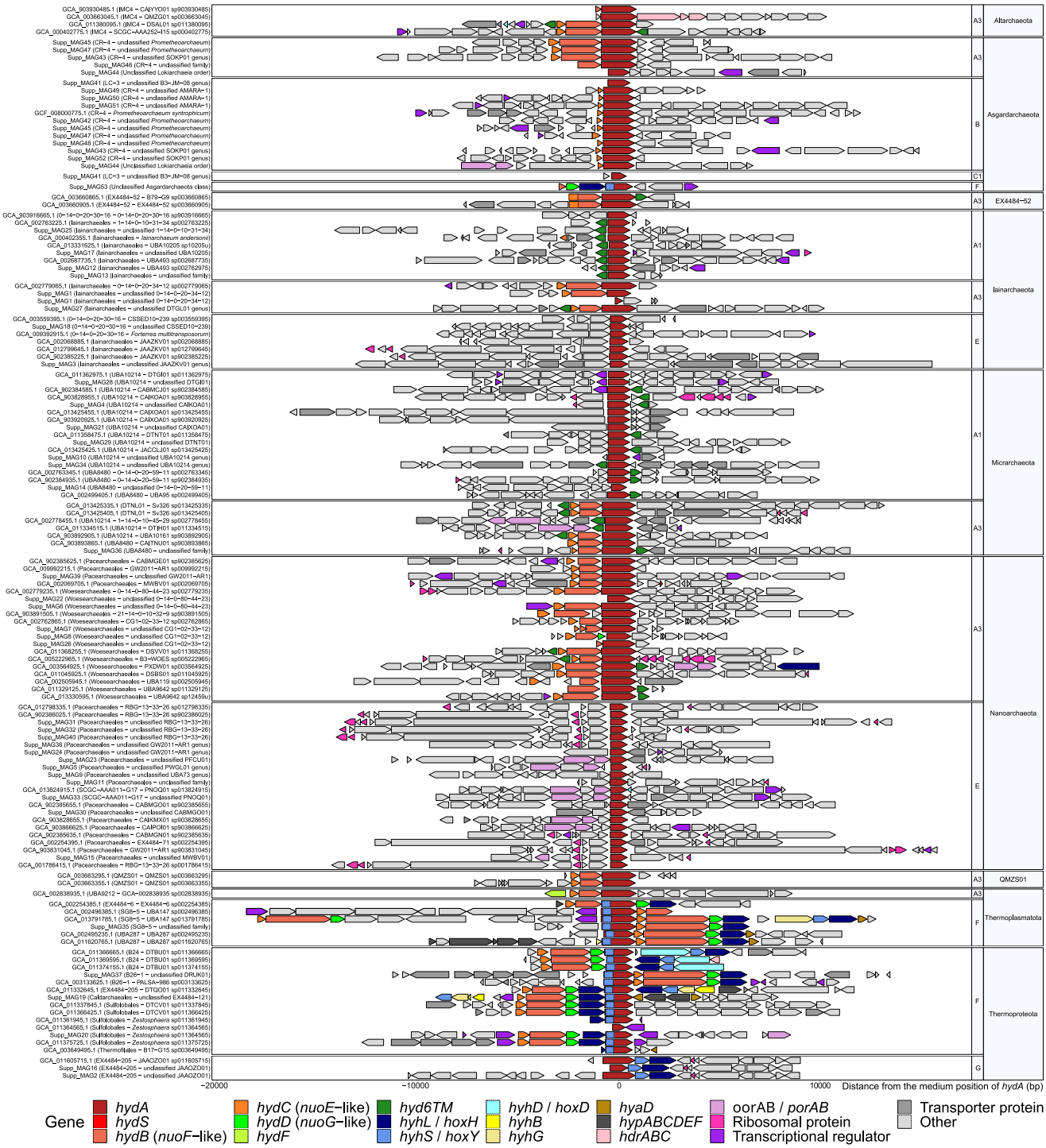


Figure S2. Genetic organization of 136 archaeal [FeFe] hydrogenases, related to Figures 1, 2, 4, and 5

Up to 10 genes upstream and downstream of the [FeFe] hydrogenase (*hydA*) are shown. Gene length is shown to scale. *hydA*, [FeFe] hydrogenase; *hydS*, [FeFe] hydrogenase small subunit; *hydB*, [FeFe] hydrogenase diaphorase subunit; *hydC*, [FeFe] hydrogenase thioredoxin subunit; *hydD*, [FeFe] hydrogenase *nuoG*-like conduit protein; *hydF*, [FeFe] hydrogenase H-cluster maturation GTPase; *hyd6TM*, uncharacterized 4 to 6-helix transmembrane protein associated with group A [FeFe] hydrogenases; *hyhL/hoxH*, group 3 [NiFe] hydrogenase catalytic subunit; *hyhS/hoxY*, group 3 [NiFe] hydrogenase small subunit; *hyhD/hoxD*, group 3 [NiFe] hydrogenase iron-sulfur subunit D; *hyhB*, group 3 [NiFe] hydrogenase diaphorase electron transfer subunit; *hyhG*, group 3 [NiFe] hydrogenase diaphorase catalytic subunit; *hyaD*, [NiFe] hydrogenase maturation protease; *hypABCDEF*, [NiFe] hydrogenase maturation factors; *hdrABC*, CoB-CoM heterodisulfide reductase subunits; *oorAB/porAB*, 2-oxoacid:acceptor or pyruvate:ferredoxin oxidoreductase alpha and beta subunits. Detailed information on loci, annotations, and amino acid sequences of each gene are available in Table S2.

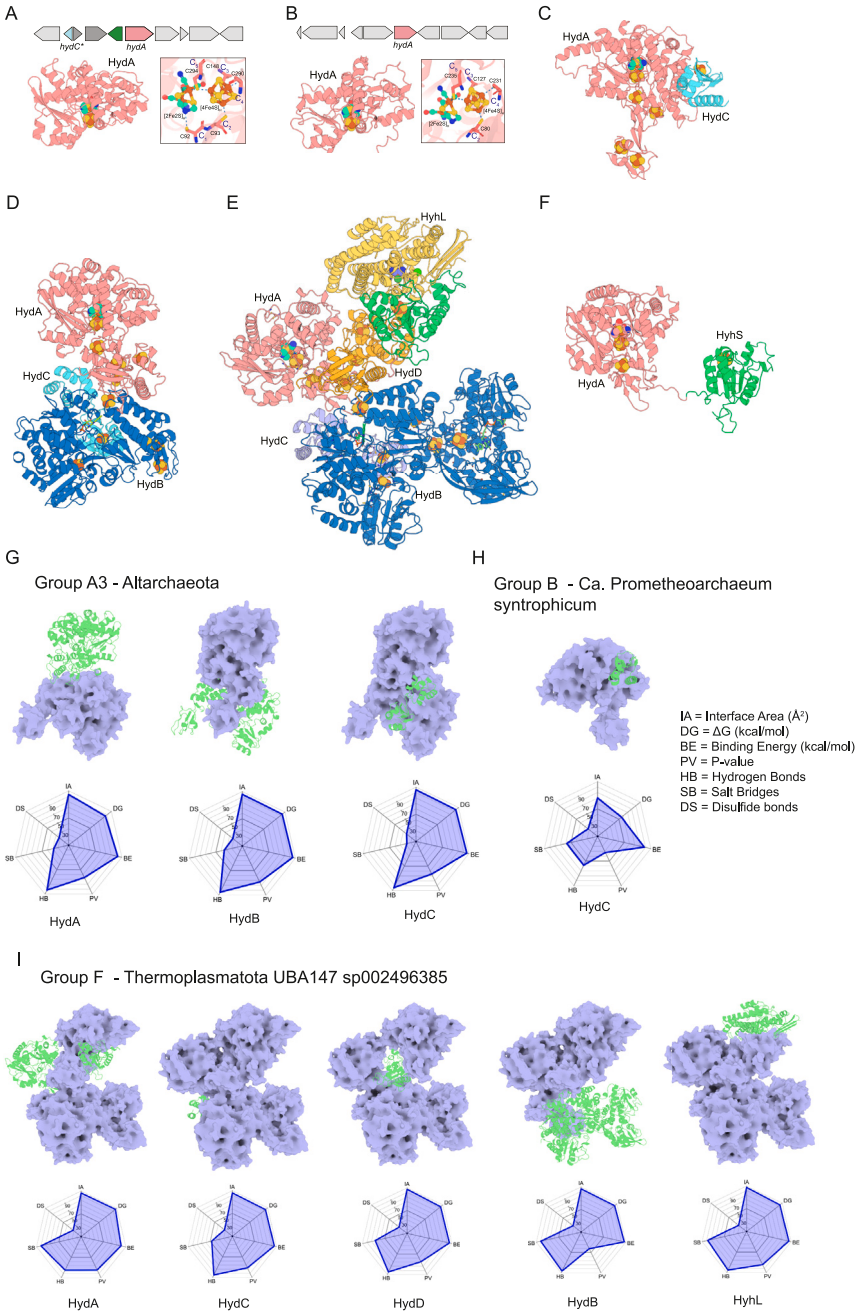


Figure S3. Analysis of AlphaFold2 models of archaeal [FeFe] hydrogenases, related to Figures 1, 2, 4, and 5

(A) Genetic organization and model of the group A1 [FeFe] hydrogenase from *Ca.* Iainarchaeum andersonii.

(B) Genetic organization and model of the group E [FeFe] hydrogenase from CABMGN01 sp902385635 (Nanoarchaeota).

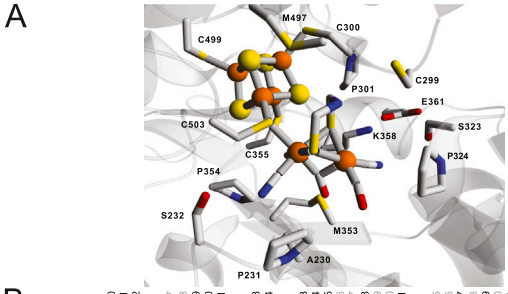
(C) Model of the group B [FeFe] hydrogenase complex from *Ca.* Prometheoarchaeum syntrophicum.

(D) Model of the group A3 [FeFe] hydrogenase complex from DSAL01 sp011380095 (Altarchaeota).

(E) Model of the complete group F [FeFe] hydrogenase from Thermoplasmatota UBA147.

(F) Model of the HydA-HyhS fusion from the group F [FeFe] hydrogenase from Thermoplasmatota UBA147.

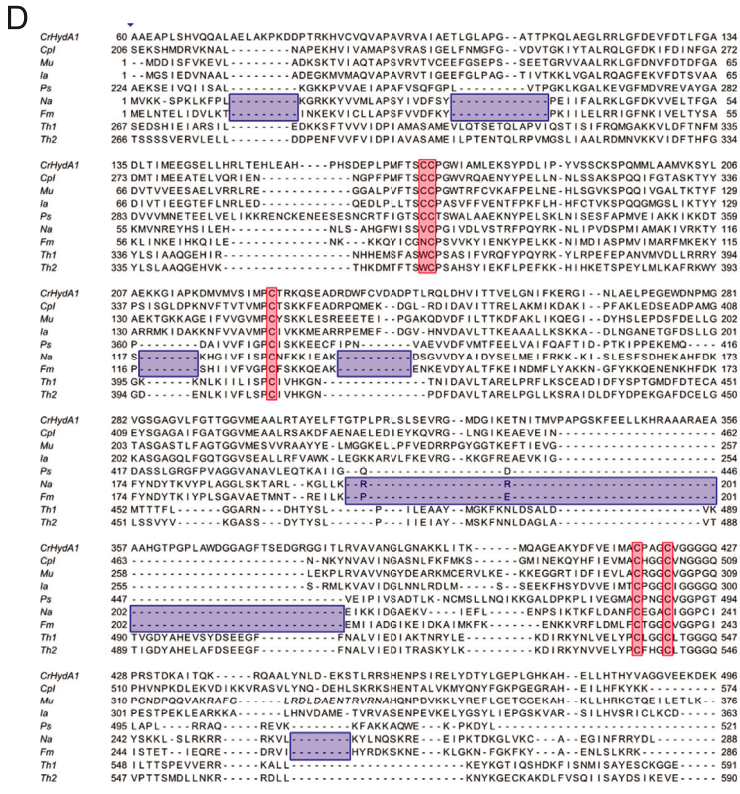
(G–I) The stability of protein-protein interfaces predicted by AlphaFold2 for each subunit of the group A3 Altarchaeota hydrogenase (G), the group B *Ca.* P. syntrophicum hydrogenase (H), and the group F Thermoplasmatota hydrogenase (I) shown with the relevant subunit displayed as a cartoon, with the rest of the complex shown as a surface view (top). PISA software package predicted parameters indicating the stability of each interface shown as a bulls-eye plot, with a larger blue area indicating that the complex is stable (bottom).



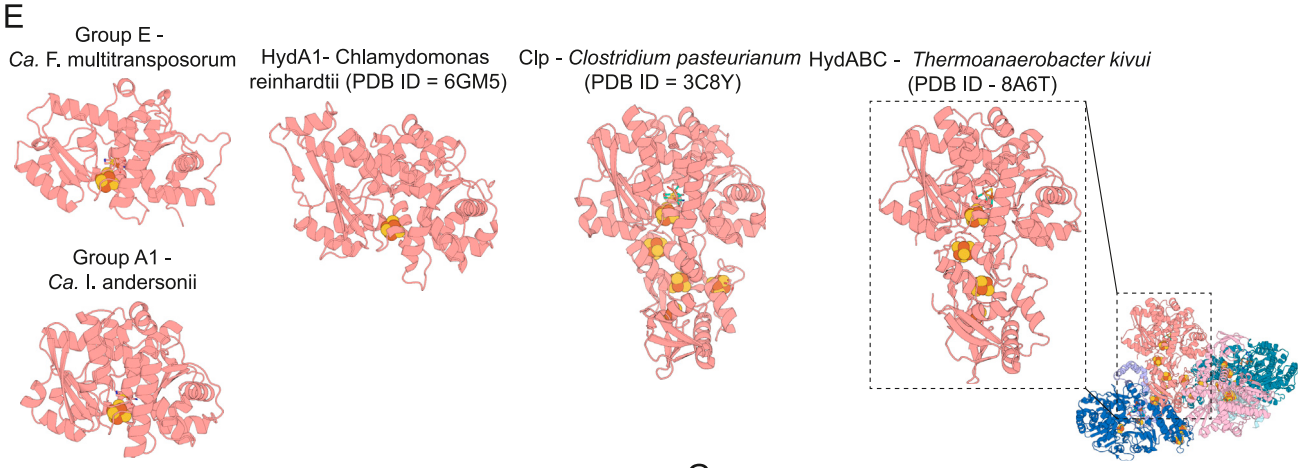


C

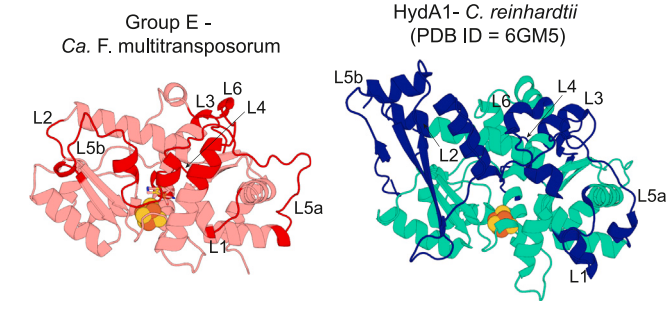
Mu (A$_{Arc}$)	APS	TSCCP	SP	MPCYSKKLE	EVLACRGGC
Ia (A$_{Arc}$)	APA	TSCCP	SP	MPCIVKKME	EIMTCPGGC
Ps (B$_{Arc}$)	APA	TSCCT	AP	GPCISKKEE	EGMACPNGC
Na (E$_{Arc}$)	APS	SSVCP	SP	SPCNFKKIE	DANFCEGAC
Fm (E$_{Arc}$)	APS	CGNCP	SP	GPCFSKKQE	DMLFCTGGC
Th1 (F$_{Arc}$)	DPI	ASWCP	EP	SPCIVHK--	ELYPCLGGC
Th2 (F$_{Arc}$)	DPI	TSWCP	SP	SPCIVHK--	ELYPCFHGC



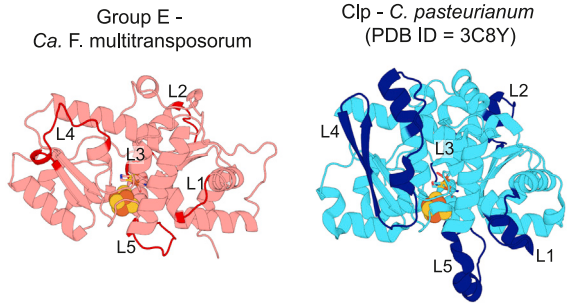
E



F



G



(legend on next page)

Figure S4. Conservation of active site residues in different classes of [FeFe] hydrogenases, sequence comparison of archaeal [FeFe] hydrogenases against bacterial [FeFe] hydrogenases, and structural comparison of the ultraminimal group E [FeFe] hydrogenase Fm, related to Figures 2 and 4

(A) Structural view of the active site of *Clostridium pasteurianum* [FeFe] hydrogenase (*Cp*I, PDB: 4XDC) showing the H-cluster and interacting amino acid residues.

(B) Normalized consensus logos of [FeFe] hydrogenase groups A–F generated in Jalview using a ClustalΩ sequence alignment of sequences retrieved from Greening et al.[1] and this work. Coloring is based on the Clustal X color scheme. Numbering is based on *Cp*I and black numbers are illustrated in the top panel.

(C) Amino acid sequences of the archaeal [FeFe] hydrogenases heterologously expressed in this work.

(D) A sequence alignment of archaeal [FeFe] hydrogenases expressed in this study with the prototypical [FeFe] hydrogenases from *Chlamydomonas reinhardtii* (CrHydA1) and *Clostridium pasteurianum* (CpI). Positions corresponding to the catalytic cysteines are highlighted in red. Gaps in the group E [FeFe] hydrogenases (Na and Fm) that contribute to the small size of this group are highlighted in blue.

(E) Cartoon representations of archaeal minimal group E and A1 [FeFe] hydrogenases, and structurally characterized minimal, monomeric, and multimeric [FeFe] hydrogenases from bacteria.

(F) A comparison of Fm with HydA1 from *C. reinhardtii* (PDB: 6GM5) with the location of connecting loops or structural elements with reduced size in Fm labeled L1 to L6 to aid comparison.

(G) A comparison of Fm with Clp from *C. pasteurianum*, displayed and labeled as in (F).

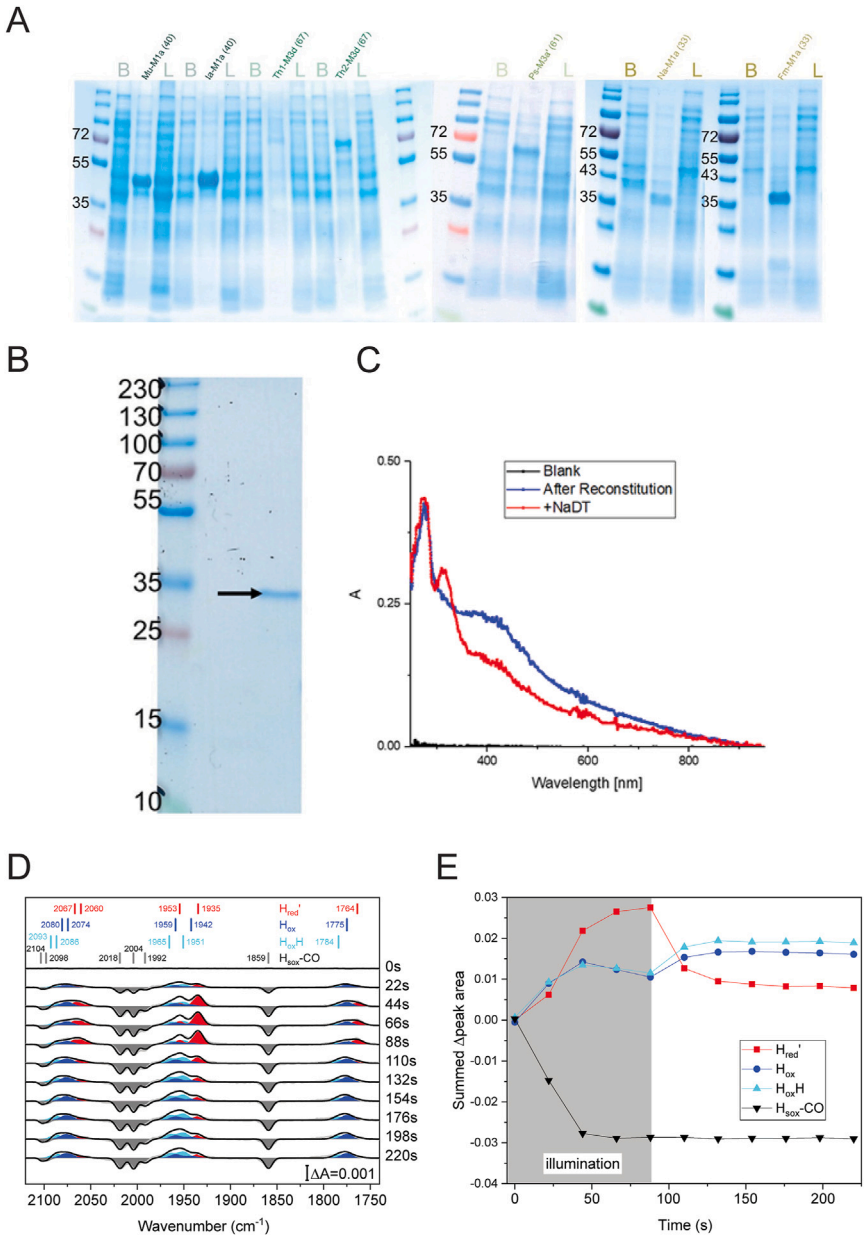


Figure S5. Heterologous expression of archaeal [FeFe] hydrogenases, and isolation, reconstitution of the [4Fe-4S]$^{+}$ cluster, FTIR difference spectra, and redox state kinetics of Fm, related to Figure 3

(A) Expression constructs with verified sequences were transformed in chemically competent *E. coli* BL21(DE3). Protein bands are shown from before induction with IPTG in lane B, after induction (name-subclass and with the expected kDa size in parenthesis), and lysate or supernatant after cell lysis and centrifugation in lane L. The bands in each after-induction lane corresponded well with the expected molecular weights in kDa. Both group A1 [FeFe] hydrogenases (Mu and Ia) had the highest expression and solubility levels. In contrast, the group F (Th1, Th2) and group B (Ps) enzymes exhibited moderate expression levels but poor solubilities in aqueous solutions. The group E (Na and Fm) enzymes exhibited high expression levels but poor solubility.

(B) SDS-PAGE gel of purified Fm (33 kDa), obtained following expression in *E. coli* Origami B(DE3) and StrepTrap XT purification. Target protein indicated with horizontal arrow.

(C) UV-visible spectra of Fm (205 μM) after reconstitution of [4Fe-4S]$^{2+}$ cluster (blue spectrum), indicated by the absorbance at 405 nm. The Fe/protein content was 4.2 ± 0.4 after reconstitution, in agreement with the presence of a single [4Fe-4S] cluster. Upon addition of 20× excess sodium dithionite (NaDT, red spectrum), a decrease in absorbance at 405 nm is observable, indicating the reduction of [4Fe-4S]$^{2+}$ cluster to [4Fe-4S]$^{+}$. Spectra were collected in a 1-mm path length cuvette.

(D) The full set of difference spectra of the photoreduction experiment shown in Figure 3B in the main text. The super-oxidized CO-inhibited H_{sox}-CO species (gray bands) depopulates in favor of the one-electron-reduced oxidized states $H_{ox}H$ (cyan bands) and H_{ox} (blue bands). During continuous photoreduction, the oxidized species are further reduced to the [4Fe4S] cluster reduced state H_{red}' (red bands). Illumination that facilitates photoreduction was applied for 88 s.

(legend continued on next page)

(E) The summed delta peak area of each redox state of the difference spectra in (D) is plotted over the time course of the photoreduction experiment. The illumination period is indicated by the gray area. The depopulation of the super-oxidized CO-inhibited H_{sox}-CO species (black) is mostly completed within 44 s. At the same time the oxidized states $H_{ox}H$ (cyan) and H_{ox} (blue) reach their maximum population during photoreduction. Subsequently during the illumination period H_{red}' accumulates at the expense of the oxidized species (until 88 s). After photoreduction (88 s) H_{red}' converts back into the oxidized species. Note that no re-population of H_{sox}-CO was detected.

CellPress
OPEN ACCESS

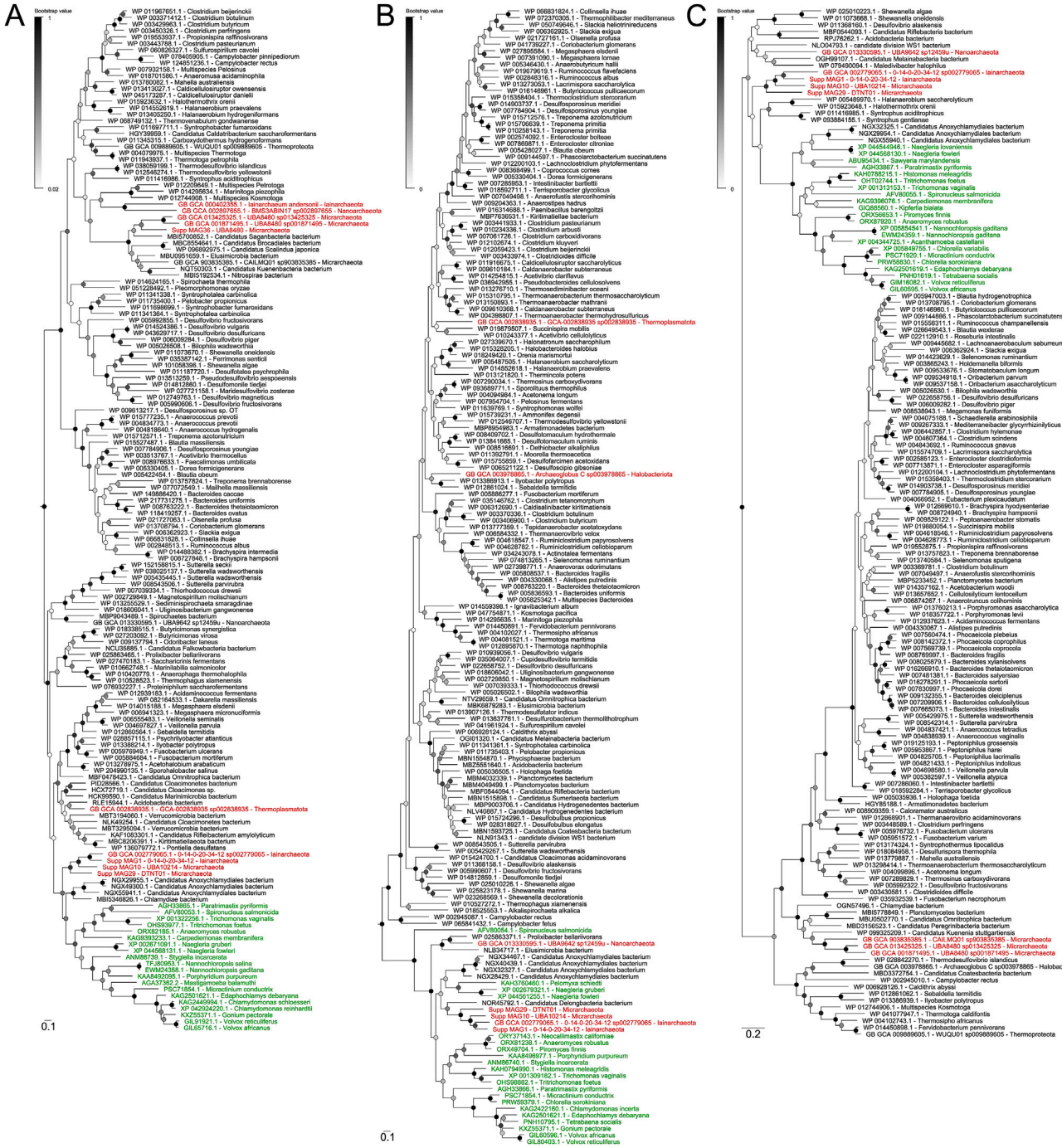


Figure S6. Phylogenetic trees of the three [FeFe] hydrogenase maturases (HydEFG), related to Figure 1

(A) HydE.

(B) HydF.

(C) HydG.

Different colors show archaeal (red), bacterial (black), and eukaryotic (green) sequences. Evolutionary history was inferred using the maximum-likelihood method and Jones-Taylor-Thornton (JTT) matrix-based model with 50 bootstrap replicates and midpoint rooting.

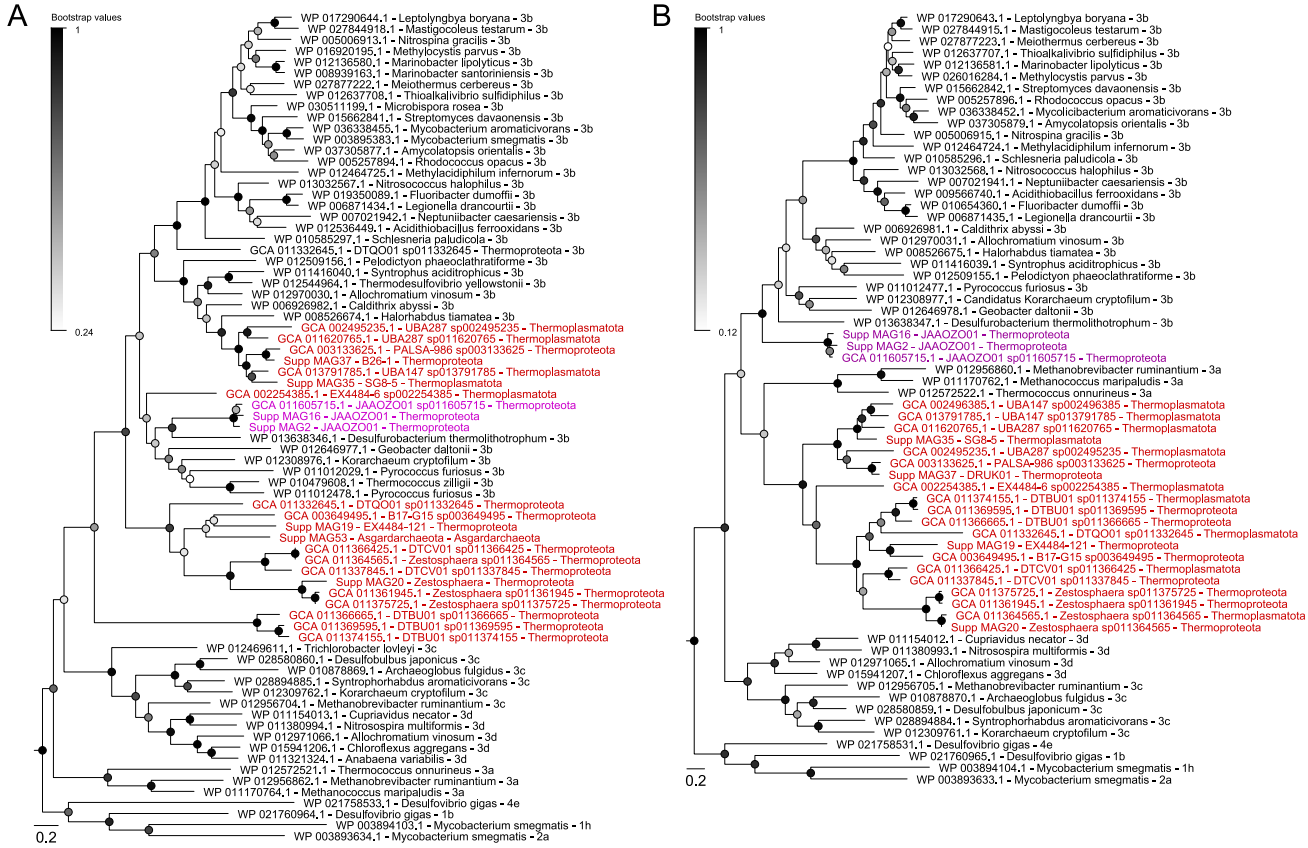


Figure S7. Phylogenetic trees of the [NiFe] hydrogenase large/catalytic subunit (HyhL) and iron-sulfur domain/small subunit (HyhS), with focus on group 3 [NiFe] hydrogenases, related to Figure 4

(A) HyhL. The subunits predicted to associate with [FeFe] hydrogenases are shown in red (for group F [FeFe] hydrogenases) and purple (for group G [FeFe] hydrogenases).

(B) The domain that fuses with the group F [FeFe] hydrogenases is shown in red. The unfused subunit encoded downstream of group G [FeFe] hydrogenases is shown in purple.

Evolutionary history was inferred using the maximum-likelihood method and Jones-Taylor-Thornton (JTT) matrix-based model with 50 bootstrap replicates and midpoint rooting.