



Diversification of methanogens into hyperalkaline serpentinizing environments through adaptations to minimize oxidant limitation

Elizabeth M. Fones¹ · Daniel R. Colman¹ · Emily A. Kraus² · Ramunas Stepanauskas³ · Alexis S. Templeton⁴ · John R. Spear² · Eric S. Boyd¹

Received: 28 May 2020 / Revised: 28 October 2020 / Accepted: 11 November 2020 / Published online: 30 November 2020
© The Author(s) 2020. This article is published with open access

Abstract

Metagenome assembled genomes (MAGs) and single amplified genomes (SAGs) affiliated with two distinct *Methanobacterium* lineages were recovered from subsurface fracture waters of the Samail Ophiolite, Sultanate of Oman. Lineage Type I was abundant in waters with circumneutral pH, whereas lineage Type II was abundant in hydrogen rich, hyperalkaline waters. Type I encoded proteins to couple hydrogen oxidation to CO₂ reduction, typical of hydrogenotrophic methanogens. Surprisingly, Type II, which branched from the Type I lineage, lacked homologs of two key oxidative [NiFe]-hydrogenases. These functions were presumably replaced by formate dehydrogenases that oxidize formate to yield reductant and cytoplasmic CO₂ via a pathway that was unique among characterized Methanobacteria, allowing cells to overcome CO₂/oxidant limitation in high pH waters. This prediction was supported by microcosm-based radiotracer experiments that showed significant biological methane generation from formate, but not bicarbonate, in waters where the Type II lineage was detected in highest relative abundance. Phylogenetic analyses and variability in gene content suggested that recent and ongoing diversification of the Type II lineage was enabled by gene transfer, loss, and transposition. These data indicate that selection imposed by CO₂/oxidant availability drove recent methanogen diversification into hyperalkaline waters that are heavily impacted by serpentinization.

Introduction

Hydrogen (H₂) links the geosphere to the biosphere and has likely done so since early in Earth's history [1–4]. H₂ can be generated through hydration and oxidation of ferromagnesian minerals (e.g., olivine) in mafic and ultramafic rocks

during serpentinization [5], a geochemical process that has been proposed to have supported life on early Earth [6, 7]. High concentrations of H₂, in turn, can abiotically reduce dissolved inorganic carbon (DIC) to generate reduced carbon substrates including carbon monoxide (CO), formate (HCOO⁻), and methane (CH₄) [8, 9], additional potent reductants capable of fueling microbial metabolism [10–19]. However, the process of serpentinization also generates hyperalkaline waters that are highly reduced and limited in DIC and other oxidants capable of supporting autotrophs, including methanogens and acetogens [19–21]. As such, the process of serpentinization can create competing benefits (replete reductant supply) and challenges (limited DIC and oxidant supply) for autotrophic subsurface microbial life, both today and early in Earth's history.

Microbial communities that inhabit hyperalkaline waters in serpentinizing environments tend to be less diverse and less abundant than those inhabiting waters with circumneutral pH [10, 18, 19, 22–24], consistent with the polyextremophilic and DIC/oxidant limited conditions commonly found at pH > 10. More specifically, the decrease in diversity among communities that inhabit hyperalkaline

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00838-1>) contains supplementary material, which is available to authorized users.

✉ Eric S. Boyd
eboyd@montana.edu

- ¹ Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA
- ² Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO 80401, USA
- ³ Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA
- ⁴ Department of Geological Sciences, University of Colorado, Boulder, CO 80309, USA

waters tends to be associated with a shift toward a predominance of strict anaerobes, a finding that is in line with the highly reduced nature of these fluids [19, 22]. Among the anaerobes commonly detected in hyperalkaline serpentinized waters are methanogens affiliated with the autotrophic and hydrogenotrophic genus, *Methanobacterium* [13, 14, 18, 19, 23, 25].

Hydrogenotrophic methanogens are often advocated as having the most ancient of extant metabolisms [4, 26, 27]. There are numerous rationales for the primacy of hydrogenotrophic methanogens. First, CH₄ in fluid inclusions in ancient rocks dated to 3.46 Ga are isotopically light, consistent with active methanogenesis at this time [28], and phylogenetic data suggest that these cells were likely dependent on H₂ as the electron donor [2, 4]. Second, the ability of hydrogenotrophic methanogens to use CO₂ as the sole carbon source and electron acceptor may have alleviated a major barrier for early autotrophic life. CO₂ is presumed to have been readily available on early Earth whereas organic carbon and other potential oxidants (e.g., oxygen, nitrate, and sulfate) are likely to have been far more limiting [29]. However, modern serpentinizing environments exhibit extremely low concentrations of DIC [13, 14, 18, 19, 23, 25], and it is not yet understood how methanogens or other autotrophs could have overcome this limitation.

CH₄ in hyperalkaline, highly serpentinized waters typically exhibits isotopically heavy carbon ($\delta^{13}\text{C-CH}_4 > -40\text{‰}$) [13, 25, 30, 31], a finding that has been suggested to result from abiotic processes or biological methanogenesis under extreme DIC limitation [25, 32, 33]. Support for the latter interpretation includes experimental evidence demonstrating that biological methanogenesis under carbon-limited conditions can result in relatively high $\delta^{13}\text{C-CH}_4$ values [34] and the presence of lipid biomarkers associated with methanogenic archaea recovered from globally distributed serpentinizing environments that exhibit high ¹³C values [35, 36]. These observations point to near-complete (quantitative) consumption of bioavailable DIC by methanogens in these environments. Furthermore, cells in hyperalkaline waters in serpentinizing systems tend to preferentially assimilate carbon substrates for biomass synthesis rather than dissimilate them for energy production when compared to cells from waters with circumneutral pH [18]. This is consistent with cells, including autotrophic methanogens, experiencing DIC limitation in hyperalkaline waters [18, 37]. Here, we apply metagenomics and single cell genomics (SCG) to examine how autotrophic methanogens overcome DIC/oxidant limitation imposed by highly reducing, high pH conditions in the Samail Ophiolite, Sultanate of Oman, the largest ophiolite and one that is undergoing active low temperature serpentinization [20, 25, 38–42]. Microcosm assays utilizing

radioisotopically labeled substrates are used to evaluate hypotheses generated from metabolic reconstructions inferred from genomic data. Finally, phylogenomic and SCG approaches are used to describe the diversification of methanogens in hyperalkaline waters resulting from serpentinization and to identify the mechanisms involved in their apparent speciation.

Materials and methods

Sample collection

A Grundfos SQ 2–85 submersible pump was used to collect water samples in February 2017 from seven previously drilled wells in the Samail Ophiolite, Sultanate of Oman, as previously described [18, 19]. Briefly, waters were collected at least 20 m below the water table in each well, including two depths at well NSHQ14: 50 m (NSHQ14B) and 85 m (NSHQ14C). After pumping ~100 L of water through the tubing and filter housing, biomass was collected using in-line 0.2 μm Millipore polycarbonate filters in 47 mm Pall polycarbonate filter housings, as described previously [18]. Filters and their contents were transferred to sterile screw top vials and were immediately flash frozen in the field with liquid nitrogen. Water samples for single cell genomics (SCG) were collected from NSHQ14C only. Filter-sterilized (0.2 μm) glycerol (5% final concentration) and TE buffer (1×) were added to samples as cryoprotectants, after which the samples were flash frozen in liquid nitrogen and stored at –80 °C until further analysis.

A Grundfos SQ 2–85 submersible pump was again used to collect water samples for methanogenic activity assays from well WAB188 at a depth of 50 m on March 6, 2020. A Double Packer Standard System (SolExperts, France) was deployed in well NSHQ14 to attempt to isolate shallow versus deep waters in the well. However, attempts to isolate deep waters were unsuccessful as there was insufficient flow to pump water to the surface when the top packer was inflated (at both 30 m and at 18 m depths). Sufficient flow was achieved when the bottom packer was inflated at a depth of 30 m with the top packer left uninflated, isolating waters from an interval extending between the depth of the water table (9 m) to 30 m on February 28, 2020. After pumping ~100 L of water through the sampling manifold and tubing at each site, water samples for methanogenic activity assays were collected by direct injection from the sampling manifold into N₂-purged, autoclave-sterilized, butyl-stoppered glass serum vials. Samples were maintained at ambient temperature during transport to the laboratory and were then placed at 4 °C for storage. Water temperature and pH were measured in the field in 2020 with a Hach (Loveland, CO) HQ40D Portable Multi Meter.

Table 1 Geochemical measurements for well waters sampled in 2017 and 2020.

Well	WAB188		NSHQ14		
	2017	2020	2017	2017	2020
Water table depth (m)	9	8.5	12	12	9
Sampled depth (m)	78	50	50	85	9–30
pH	7.6 ^a	7.5	11.1	11.3	11.3
Temperature (°C)	33.0 ^a	34.7	34.4	36.3	35.7
H ₂ (μM)	0.92	–	21	164	–
DIC (mM)	3.0	–	0.05	0.13	–
Formate (μM)	1.0 ^b	–	1.7 ^{b,c}	1.7 ^{b,c}	–
CH ₄ (μM)	1.69	–	34.6	12.6	–

– Denotes data that are not available from the 2020 field season.

Values measured in 2017 were previously reported in Fones et al. [18].

^aData collected from wells in 2016 that were previously reported in Rempfert et al. [19] are displayed since measurements were not collected during the 2017 field season.

^bData collected from wells in 2015 that were previously reported in Rempfert et al. [19] are displayed since measurements were not collected during the 2017 field season.

^cThe formate concentration was measured in only one depth interval (20 m) in NSHQ14 in 2015, as previously reported in Rempfert et al. [19].

Site description

The classification scheme used here to describe major water types (hyperalkaline peridotite, contact, and alkaline peridotite) in the Samail Ophiolite is as reported previously [18, 19]. Increasing pH is generally interpreted to reflect the extent of serpentinization reaction progress and the degree of fluid mixing in each well [19]. DIC was replete (3.0–3.5 mM) and H₂ was limited (undetectable to 0.92 μM) in circumneutral waters (pH 7.6–8.5) from wells WAB188, WAB105, and WAB104 at the time of sampling in 2017 (Table 1) [18]. In contrast, DIC was limited (0.05–0.13 mM) and H₂ was replete (21 to 164 μM) in hyperalkaline waters (pH 11.1–11.3) from well NSHQ14 in 2017. Formate (HCOO⁻), another potential reductant generated during serpentinization, has been measured at concentrations ranging from 1.0 to 1.7 μM in these wells [19]. Further details of the geochemistry of waters sampled from the Samail Ophiolite are reported elsewhere [18, 19].

DNA extraction and shotgun metagenomic sequencing

DNA was extracted for metagenomic sequencing as previously described [18]. Briefly, DNA was extracted from filtered biomass using a Zymo (Irvine, CA) Research Xpedition Soil/Fecal DNA MiniPrep Extraction kit according to manufacturer instructions. Sequencing libraries

were prepared in triplicate using the Nextera XT kit (Illumina Inc., San Diego, CA) and sequenced on the Illumina HiSeq 2500 Rapid Run platform (2 × 250 bp).

Metagenomic assembly and binning of metagenome assembled genomes (MAGs)

Raw sequence reads were quality filtered, trimmed of adapters, and assembled as described previously [18]. Contigs were binned into draft MAGs based on tetranucleotide frequencies and coverage profiles using the MetaBAT software package (v. 0.26.3) [43]. Contigs with ≥98% nucleotide identity to the *Methanobacterium* Type I MAG from WAB188 were classified as Type I and contigs with ≥98% nucleotide identity to the *Methanobacterium* Type II MAG from NSHQ14C were classified as Type II. The number of reads affiliated with *Methanobacterium* Type I or II MAGs as a percentage of the total reads in each metagenome were used to calculate estimated relative abundances of each population.

Single cell genomics (SCG)

Individual microbial cells were separated by fluorescence-activated cell sorting, lysed by a combination of freeze-thawing and KOH, and their genomes were amplified by WGA-X and subject to low-coverage sequencing and assembly at the Bigelow Laboratory Single Cell Genomics Center, as described previously [44]. MAG and single amplified genome (SAG) data have been submitted to GenBank under the BioProject number PRJNA631088.

Phylogenomic analyses

Phylogenomic reconstruction of target MAGs (i.e., *Methanobacterium* spp. described below) recovered from the Samail Ophiolite and representative reference *Methanobacteria* genomes were performed using a concatenation of 103 aligned, single-copy, phylogenetic marker genes, as previously described [45]. Briefly, all genomes classified within the ‘*Methanobacteria*’ class were retrieved from the Department of Energy-Integrated Microbial Genome (IMG) and National Center for Biotechnology Information (NCBI) GenBank databases. Marker genes were identified from the genome set using Amphora2 [46], individually aligned using Clustal Omega (v.1.2.4) [47], and concatenated. The concatenated alignment was subject to Maximum Likelihood phylogenetic reconstruction in IQ-TREE (v.1.6.11) [48] after identifying the optimal amino acid substitution model (LG + F + R10) via the Bayesian Information Criterion, as implemented in modelfinder [49].

Metabolic reconstructions of MAGs and SAGs

Gene prediction and annotation for MAGs and SAGs were performed with Prodigal v.2.6.3 [50] as implemented in Prokka v.1.11 [51] specifying default parameters. Following preliminary predictions and annotations for MAGs, gene functions were further examined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) function database [52] with the KEGG Automatic Annotation Server (KAAS) [53]. Genomes were also queried via BLASTp [54] for specific gene functions, as guided by gene contents of closely related genomes. BLASTp searches were conducted using bait sequences specific for active site subunits for each protein or protein complexes. Positive matches for protein homologs in MAGs and SAGs were considered as those with a BLASTp *E*-value of $<1 \times 10^{-6}$, $>30\%$ amino acid sequence similarity, and $>60\%$ of the length of the query sequence. The complete WAB188, NSHQ14B, and NSHQ14C metagenome assemblies were queried for genes encoding homologs of key proteins initially absent from the MAGs (i.e., H_2 -dependent methenyltetrahydromethanopterin dehydrogenase, Hmd, F_{420} -dependent methylenetetrahydromethanopterin dehydrogenase, Mtd, and the formate and formate/nitrate transporters, FdhC, FocA, and YfdC, in both Type I and Type II MAGs; the ion-translocating [NiFe]-hydrogenase complex, Mrp-Mbh, in the Type I MAG; the Group 3 F_{420} -reducing [NiFe]-hydrogenase, Frh, and Group 3c bifurcating [NiFe]-hydrogenase, Mvh, in Type II MAGs, described below) via BLASTp. Single amplified genomes affiliated with *Methanobacterium* were also queried for proteins missing from the MAGs via BLASTp. Unbinned contigs containing these homologs were manually assigned to MAGs based on consensus with contig coverage profiles and average GC content, as determined using CheckM [55].

Homologs of [NiFe]-hydrogenase were identified by BLASTp using query sequences for active site subunits. Briefly, candidate [NiFe]-hydrogenase catalytic subunits were compiled and combined with references for specific hydrogenase groups, as defined previously [56]. The combined dataset was aligned with Clustal Omega, evaluated manually for the presence of conserved N- and C-terminal CxxC motifs [57], and subjected to Maximum Likelihood phylogenetic analysis, as described above. The putative directionality and redox partner coupling, as determined initially from phylogenetic analyses above, were further evaluated by comparing amino acid conservation in the L1 and L2 motifs to previously characterized hydrogenases [56]. These functional assignments were further assessed by examining the genes co-localized with the genes encoding hydrogenase catalytic subunits for key partner proteins (e.g., F_{420} binding

proteins), as determined with the Conserved Domain Database [58] and BLASTp analysis against the NCBI non-redundant database.

Comparison of MAGs and SAGs

The average nucleotide identities (ANIs) among SAGs and MAGs were calculated using the MUMmer (v4.0.0beta2) average_nucleotide_identity.py script in python [59]. SAGs (B04, E10, J15, M05, P19, O02, and O05) that were found to have insufficient hits to calculate pairwise ANIs with at least one other genome were removed from this analysis. ANIs among the SAGs and the NSHQ14 MAGs were used to construct a dissimilarity matrix that was translated into a Euclidean distance matrix and used to construct a dendrogram using the hclust function in R specifying the “single linkage” agglomeration method [60].

To evaluate the similarity of proteins encoded among the MAGs and SAGs, the percent identities between each encoded protein homolog were first determined. Homologs were defined as those with $\geq 30\%$ identity to protein coding genes in the NSHQ14C MAG (used as a reference genome), as identified by BLASTp searches of the proteins encoded by each SAG. The amino acid identities of proteins encoded by the SAGs relative to proteins encoded in the NSHQ14C MAG were then averaged to calculate an overall average amino acid identity for each SAG.

Orthologous protein groups among the SAGs were first defined as those proteins encoded in SAGs that were most closely related to a single protein from the reference genome (NSHQ14C Type II MAG), as determined by BLASTp. Among the most commonly identified SAG orthologs that exhibited differences in pairwise amino acid identity were transposases. Transposase orthologs were further classified using ACLAME v.0.4 [61] and were then subjected to alignment with Clustal Omega and maximum-likelihood phylogenetic reconstruction with PhyML v.3.0 via the Bayesian Information Criterion as implemented in the Smart Model Selection in PhyML [62, 63]. The synteny of genes flanking transposases was assessed by aligning contigs encoding transposases using progressiveMauve [64], and alignments were visualized using Gene Graphics [65]. To evaluate the possibility that divergent orthologous proteins contributed to apparent diversification of SAGs, Mantel tests describing correlations between protein ortholog amino acid dissimilarities and whole genome ANI dissimilarities were performed for the ten most abundant orthologous proteins identified in SAGs that differ from those in the NSHQ14 Type II MAG. Mantel tests were performed using the mantel function of the vegan: community ecology package in R specifying 100 permutations using the Pearson correlation method [66].

Methanogenesis rate potentials

Potential rates of biological transformation of HCO_3^- and HCOO^- to CH_4 were measured via microcosm-based activity assays containing ^{14}C -radiolabeled substrates as previously described [18]. Briefly, well water samples stored at 4°C were homogenized and 10 mL aliquots were transferred from storage vials to 26 mL N_2 -purged, autoclave-sterilized, butyl-stoppered experimental vials (four biological replicates and four abiotic controls per condition) using a sterile, N_2 -purged needle and syringe. Abiotic controls were sterilized by autoclaving twice with incubation at room temperature overnight between autoclave cycles to allow for germination of spores. 1 mM (final concentration) of HCOO^- or HCO_3^- that included $5\ \mu\text{Ci}$ $^{14}\text{C}\text{-HCOO}^-$ or $^{14}\text{C}\text{-HCO}_3^-$ (American Radiolabeled Chemicals, St. Louis, MO), respectively, was then added to each vial. H_2 gas was added in excess of atmospheric pressure to each vial to account for 20% of headspace volume. Following incubation for 1, 2, 4, 6, and 8 weeks at in situ temperature (34°C), 3 mL filtered ($0.2\ \mu\text{m}$) N_2 was added to each vial using an N_2 -purged sterile needle and syringe to maintain positive headspace pressure, and 3 mL headspace gas was removed using a needle and syringe equipped with a stopcock. The 3 mL headspace gas from each microcosm vial was then injected into a specially fabricated scintillation vial containing a butyl rubber septum and CH_4 was trapped with 10 mL Cytoscint ES scintillation cocktail, as described previously [18]. The radioactivity, measured in counts per minute, associated with $^{14}\text{C}\text{-CH}_4$ from each of the samples was measured on a PerkinElmer Tri Carb 2900TR Liquid Scintillation Analyzer (PerkinElmer, Waltham, MA), converted to disintegrations per minute using a quench curve, normalized to account for total substrate added (radiolabeled and unlabeled) and total headspace gas volume, and used to calculate the rates of substrate transformation. $^{14}\text{C}\text{-CH}_4$ values in abiotic controls were subtracted from $^{14}\text{C}\text{-CH}_4$ values in biological replicates to arrive at the amount of $^{14}\text{C}\text{-CH}_4$ produced that could be attributed to biology. Statistical significance of differences between biological assays and abiotic controls were assessed via Student's *t*-test assuming unequal variance for each condition.

Results and discussion

Reconstructed genomes from metagenomic sequences were used to develop metabolic models for methanogen populations inhabiting waters that span a range of geochemical conditions due to differences in serpentinization reaction progress. Two phylogenetically distinct lineages affiliated with *Methanobacterium* were the sole methanogens

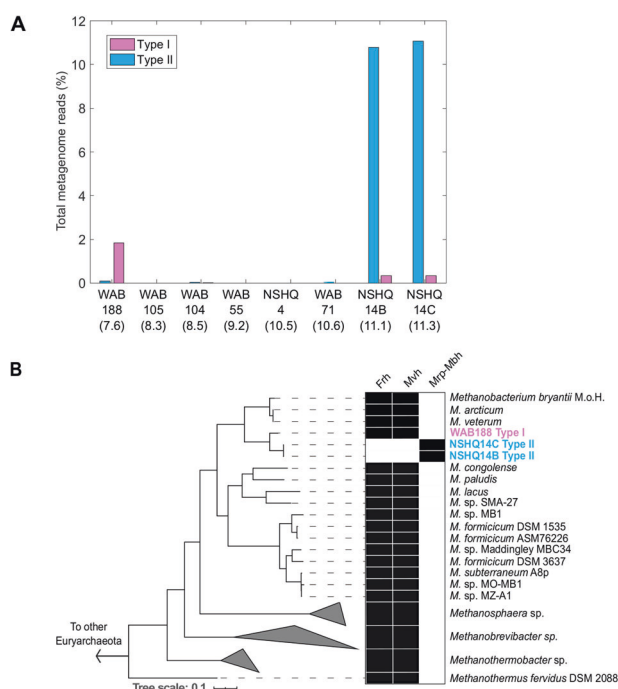


Fig. 1 The estimated relative abundances of *Methanobacterium* MAGs in communities from subsurface fracture waters collected from wells intersecting the Samail Ophiolite in 2017 (A) and phylogenomic reconstruction of Oman methanogen metagenome assembled genomes (MAGs) in relation to representative *Methanobacterium* genomes (B). In A, wells are listed in order of ascending pH from left to right with pH values indicated below the name of each well in parentheses. Contigs with $\geq 98\%$ nucleotide identity to the *Methanobacterium* Type I MAG from WAB188 were classified as Type I and contigs with $\geq 98\%$ nucleotide identity to the *Methanobacterium* Type II MAG from NSHQ14C were classified as Type II. Estimated relative abundances are shown as the number of reads affiliated with *Methanobacterium* Type I or II MAGs as a percentage of the total reads in each metagenome. In B, pink text depicts the Type I MAG and cyan text depicts Type II MAGs. All bootstrap values of displayed nodes are >980 out of 1000 bootstrap replicates. Clade-level triangles indicate the phylogenetic diversity within each group via side lengths that are proportional to the distances between the clade's most closely related and furthest related taxa. Branch length is relative to the scale provided at the bottom of the figure indicating the expected number of substitutions per site. Filled boxes to the right of terminals indicate the presence of genes in reconstructed MAGs or genomes, whereas empty boxes indicate absence of genes in reconstructed MAGs or genomes. *Frh* coenzyme F_{420} -reducing (Group 3a) [NiFe]-hydrogenase, *Mvh* methyl viologen (Group 3c) [NiFe]-hydrogenase, *Mrp-Mbh* multiple resistance and pH adaptation module—membrane-bound (Group 4) [NiFe]-hydrogenase protein complex.

recovered from subsurface well waters from the Samail Ophiolite (Fig. 1). One MAG (Type I) represented an abundant microbial lineage in a well with circumneutral water (WAB188: pH 7.6), whereas another MAG (Type II) represented a lineage that was abundant in a well with hyperalkaline water, including samples collected at both 50 m (NSHQ14B: pH 11.1) and 85 m depths (NSHQ14C: pH 11.3) (Fig. 1A). MAGs affiliated with the Type II lineage from NSHQ14B and from NSHQ14C exhibited

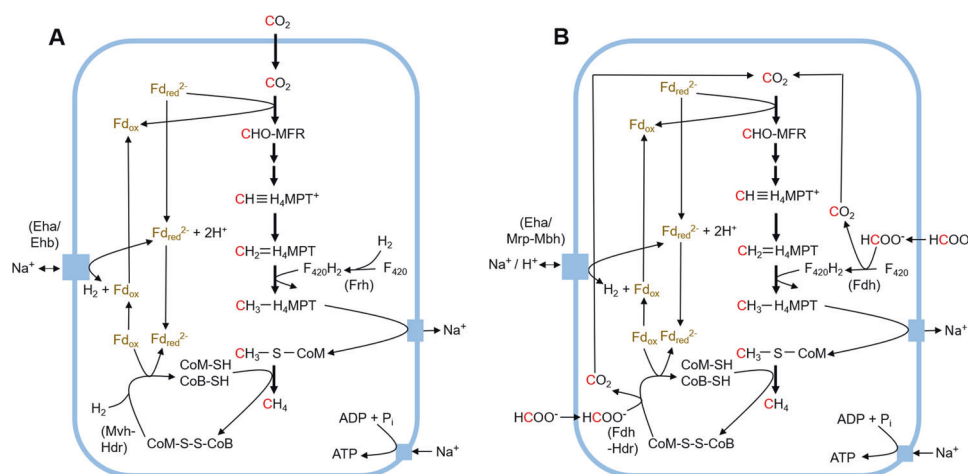


Fig. 2 Proposed hydrogenotrophic and formatrophic methanogenesis pathways in *Methanobacterium* Type I (A) and Type II (B) populations, respectively, from the Samail Ophiolite, Oman. Enzymes that differed between the two MAGs are denoted in the figures. *Eha* Group 4d energy-converting [NiFe]-hydrogenase A, *Ehb* Group 4d energy-converting [NiFe]-hydrogenase B, *Fd* ferredoxin,

Fdh formate dehydrogenase, *Frh* coenzyme F_{420} -reducing (Group 3a) [NiFe]-hydrogenase, *Hdr* heterodisulfide reductase, *MFR* methanofuran, *MPT* methanopterin, *Mrp-Mbh* multiple resistance and pH adaptation module—membrane bound (Group 4) [NiFe]-hydrogenase protein complex, *Mvh* methyl viologen-reducing (Group 3c) [NiFe]-hydrogenase. Figure adapted from Boyd et al. [1].

>99.99% identity to each other on an average nucleotide level (data not shown). Due to the extremely high similarity between these two MAGs, one MAG (*Methanobacterium* Type II from NSHQ14C) was selected as the reference genome for the Type II lineage. Contigs affiliated with both Type I and Type II lineages were also detected at low abundances in most of the other wells, which is consistent with the previous finding that 16S rRNA genes affiliated with *Methanobacterium* were widely distributed at varying abundances among subsurface well waters collected from the Samail Ophiolite (e.g., previously detected in 16 of 20 samples in relative abundances of <3% to 28.0% of total 16S rRNA gene transcripts) [19]. Accordingly, sequences closely affiliated with *Methanobacterium* have been detected in a range of other serpentinites and often in high abundance [13, 14, 23], suggesting that this genus is a ubiquitous member of globally distributed serpentinite communities.

All *Methanobacterium* MAGs exhibited high levels of estimated completeness and low contamination (Supplementary Table S1), and were classified as high-quality draft genomes [67]. The Type I *Methanobacterium* MAG was larger (1.904 Mbp) than the NSHQ14C *Methanobacterium* Type II MAG (1.511 Mbp) and encoded more proteins (2052 versus 1806), despite both MAGs exhibiting the same estimated levels of completeness (98.4%) (Supplementary Table S1). This finding is consistent with previous results that revealed an inverse relationship between the average estimated genome size and the pH of subsurface well waters in the Samail Ophiolite [18]. This observation is also consistent with data from the serpentinization impacted environment termed The Cedars (California, USA), where

members of a microbial community from a hyperalkaline spring (pH ~12) harbored the smallest genomes reported for their respective taxa [68]. Microorganisms residing in highly serpentinized waters presumably exhibit streamlined genomes to minimize energetic costs and nutrient demands associated with their replication and repair under extreme conditions, as has been suggested previously [18, 69].

Phylogenomic reconstruction of Oman methanogen MAGs in relation to other *Methanobacterium* genomes suggested that the taxa corresponding to Type I and II MAGs diverged relatively recently within the *Methanobacterium* lineage (Fig. 1B). The Type I and II MAGs were most closely related to each other, and unexpectedly, the Type II MAG appeared to correspond to a taxon that was descended from the Type I taxon, as described below. The Type I MAG encoded homologs of two key [NiFe]-hydrogenases, Group 3c bifurcating [NiFe]-hydrogenase (*Mvh*) and Group 3a F_{420} -reducing [NiFe]-hydrogenase (*Frh*) (discussed below) (Figs. 1B; S1), which are common to other *Methanobacteria*. However, homologs of these genes were not detected in any of the Type II genomes, including both Type II MAGs and all single amplified genomes (SAGs, described below). Furthermore, the large subunits of Group 3 [NiFe]-hydrogenases encoded by the Type I MAG were used to query the entire assemblies of both NSHQ14 metagenomes with no positive matches identified. Together, these data indicated gene loss in the Type II lineage. In contrast, among available *Methanobacteria* genomes, only the Type II lineage encoded a homolog of the ion-translocating [NiFe]-hydrogenase complex, *Mrp-Mbh*, indicating gene acquisition (Figs. 1B; S2). The large subunit of the *Mrp-Mbh* [NiFe]-hydrogenase encoded by the Type II MAGs was not detected in the

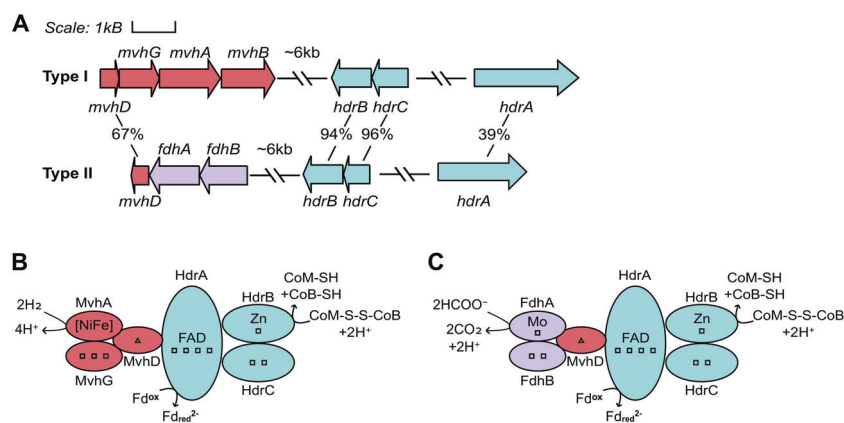


Fig. 3 Genes inferred to be co-localized with those coding for methyl viologen-reducing (Group 3c) [NiFe]-hydrogenase (Mvh) and formate dehydrogenase (Fdh) in Type I and II MAGs (A) and the putative protein complexes they form to bifurcate electrons from H_2 (B) or formate (C), respectively, to simultaneously reduce ferredoxin and heterodisulfide. Percentages below each gene indicate amino acid identities between homologs encoded by Type I and

Type II MAGs. In panels B and C, squares represent [4Fe-4S] clusters and triangles represent [2Fe-2S] clusters. CoB coenzyme B, CoM coenzyme M, FAD flavin adenine dinucleotide, Fd ferredoxin, Fdh formate dehydrogenase, Hdr heterodisulfide reductase, Mvh methyl viologen-reducing (Group 3c) hydrogenase. Figure modified from Thauer et al. [71] and Costa et al. [77].

WAB188 metagenomic assembly. Phylogenetic analysis of the large subunit of the Mrp-Mbh [NiFe]-hydrogenase encoded by the Type II MAGs revealed it to be nested among bacterial Group 4d [NiFe]-hydrogenase homologs, suggesting acquisition via horizontal gene transfer from an ancestor of the Firmicutes or Thermotogae (Supplementary Table S2; Supplementary Fig. S2). Members of the Firmicutes are commonly identified in subsurface waters of the Samail Ophiolite [19] and other serpentinizing systems [14, 23], leading to the possibility that the genes encoding Mrp-Mbh could have been acquired in a similar environment type. As such, the divergence between Type I and Type II MAGs appears to have occurred relatively recently, a conclusion that is supported by short estimated branch lengths separating these lineages (Fig. 1B).

Metabolic reconstructions of *Methanobacterium* Type I and II MAGs indicated that both organisms encoded methanogenesis pathways, albeit with several key differences as mentioned briefly above (Fig. 2). The Type I MAG encoded all requisite genes for hydrogenotrophic methanogenesis in *Methanobacteria* [70] (Fig. 2A), including Group 4 energy-converting [NiFe]-hydrogenase (Eha/Ehb), Frh, and the Mvh-heterodisulfide reductase complex (Mvh-Hdr) (Supplementary Figs. S1; S2). Homologs of H_2 -dependent methenyltetrahydromethanopterin dehydrogenase (Hmd) and F_{420} -dependent methylenetetrahydromethanopterin dehydrogenase (Mtd), which can function independently or in tandem to catalyze a critical step in hydrogenotrophic methanogenesis [71, 72], were initially not detected in either the Type I or Type II MAG. However, contigs containing homologs of Mtd with 89% identities to a previously described

Methanobacterium sp. per BLASTp were later identified in unbinned contigs of the WAB188 metagenome and both NSHQ14 metagenomes and were assigned to the Type I and Type II MAGs, respectively (Supplementary Table S3).

The Type II MAG encoded Eha (Supplementary Figs. S2; S3) but surprisingly did not encode Frh (Fig. 1B; Supplementary Table S4), the latter of which couples the oxidation of H_2 to the reduction of coenzyme F_{420} [56, 73]. The Type II MAG also did not encode Mvh (Fig. 1B; Supplementary Table S4), which couples with Hdr to bifurcate electrons from H_2 to simultaneously reduce low potential ferredoxin (Fd) and the heterodisulfide bond between coenzyme M and coenzyme B (CoM-S-S-CoB) [56]. Mvh-Hdr and Frh are [NiFe]-hydrogenases characteristically encoded in the genomes of all other members of the *Methanobacteria* (Fig. 1B; Supplementary Table S4) and are common features associated with all hydrogenotrophic (Class I) methanogens [70]. Importantly, genetic experiments indicated that neither Mvh nor Frh homologs are required during methanogenic growth with formate [74, 75]. For example, in *Methanococcus maripaludis* Vhu (alternative name for Mvh) and Frh deletion strains, these functions are thought to be replaced by a formate dehydrogenase (Fdh)–Hdr complex and a F_{420} -reducing Fdh, respectively [74–78]. FdhAB and VhuD proteins have been suggested to form a complex with HdrABC that allows for bifurcation of electrons from formate and their coupled reduction of heterodisulfide and Fd [76, 79]. In support of this, the alpha subunit of Fdh from *M. maripaludis* co-purified with the beta subunit of Fdh, the delta subunit of Vhu, and all subunits (i.e., ABC) of Hdr

[76, 79]. In addition, the tungsten-containing formylmethanofuran dehydrogenase, which catalyzes the reduction of CO₂ and covalent attachment of methanofuran in the first step of methanogenesis, was found to co-purify with FdhAB, HdrABC, and VhuD in *M. maripaludis* when grown under formatotrophic, but not hydrogenotrophic, conditions [76]. This may help to explain how *M. maripaludis*, and potentially other hydrogenotrophic methanogens, prevent the loss of cytoplasmic CO₂ for use in methanogenesis and biomass generation during formate-dependent growth.

The biochemical and phenotypic observations of Frh and Vhu deletion strains of *M. maripaludis* provide a framework to rationalize how the Type II cells from the Samail Ophiolite, which lack Frh and Mvh, might catalyze methanogenesis. The Type II MAG encoded Fdh (*fdhAB*) and this was inferred to be co-localized in the genome with *mvhD* and *hdrBC* (Fig. 3). While the structure of F₄₂₀-reducing Fdh has not yet been resolved [77], Fdh purified from *Methanobacterium formicum* reduced coenzyme F₄₂₀ [80, 81], and the beta subunit (FdhB) was predicted to have an F₄₂₀-binding domain similar to that of the Frh beta subunit (FrhB) [82]. The residues involved in F₄₂₀ coordination by FrhB in *Methanothermobacter marburgensis* (positions 163, 165–167, and 208–211) [82] were conserved in FdhB proteins from both *M. formicum* and the Type II *Methanobacterium* MAG with two exceptions (Supplementary Fig. S4). The two positions that were not conserved in *M. formicum* and the Type II MAG, however, harbor substitutions to amino acids of the same polarity and charge (S to N and V to W) and were consistent between both FdhB proteins. The mechanism of formate transport into the cell remains somewhat unclear as *Methanobacterium* Type I and Type II MAGs both lacked homologs of the canonical formate transporters FdhC, YfdC, and FocA. Each *Methanobacterium*-affiliated SAG and the entire metagenomic assemblies from WAB188, NSHQ14B, and NSHQ14C were queried for these possible formate transporters with no positive matches identified. However, the Type II MAG encoded a homolog (38.8% identity, 95% query coverage) of the formate:oxalate antiporter, OxIT (Supplementary Table S2), which was lacking from the Type I MAG and may function in transporting formate into the Type II cells. These data collectively suggest that the Type II *Methanobacterium* may drive methanogenesis via a pathway that is unique among characterized Methanobacteria, wherein formate is required as electron donor and with H₂ serving as a source of additional reductant (Fig. 2B).

Consistent with previous work suggesting an essential anaerobic role for Eha during formatotrophic methanogenesis [74], both the Type I and II MAGs encoded homologs of this enzyme complex (Supplementary

Figs. S2, S3). The Type II MAGs, but not Type I MAG or other members of the Methanobacteria (Fig. 1B), also encoded a [NiFe]-hydrogenase that is homologous to the membrane-bound Mrp-Mbh complex in *Pyrococcus furiosus* (Supplementary Fig. S5) [83]. Mrp-Mbh in *P. furiosus* comprises 14 subunits, including a Na⁺/H⁺ antiporter domain (Mrp) and a [NiFe]-hydrogenase domain (Mbh) [83]. In *P. furiosus*, the Mrp-Mbh complex catalyzes the reversible oxidation of Fd coupled with H₂ production, with excess potential used to pump Na⁺ or H⁺ outside of the cell. In this case, oxidation of Fd would generate an electrochemical gradient that could be used to drive ATP synthesis. Alternatively, Fd reduction could be coupled to H₂ oxidation, with the shortage of potential compensated for by releasing Na⁺ or H⁺ into the cell. Consequently, the Type II *Methanobacterium* cells could use this complex to: (1) generate reduced Fd and (2) neutralize cytoplasmic pH (if the coupling ion is H⁺). Indeed, Mrp was first characterized in the alkaliphile *Bacillus halodurans*, where it was found to be critical for pH homeostasis under alkaline conditions [84]. The reversibility of the enzyme system may allow for these functions to shift with changing cellular demands.

Why would *Methanobacterium* Type II have evolved to drive methanogenesis with formate rather than H₂ as primary reductant in an environment such as NSHQ14, where H₂ concentrations (21–2900 μM) were over 1–3 orders of magnitude higher than formate (1.7 μM) (Table 1) [18, 19, 25]? The standard state reduction potentials of hydrogen and formate are similar (−414 and −420, respectively) [85], although the nonstandard state potential of H₂ is likely to be lower than that of formate in environments where the concentration of H₂ exceeds formate. It is also possible that the differential mobility of formate and H₂ in aqueous solutions could favor formate utilization [86]. Alternatively, we propose adaptation to extreme CO₂ limitation in highly serpentinized waters as a likely explanation. Although H₂ was replete in the environment inhabited by Type II *Methanobacterium* cells, use of H₂ as sole electron donor for methanogenesis necessitates coupling with CO₂ as both electron acceptor and source of carbon [70]. However, dissolved CO₂ is extremely limited in highly serpentinized waters due to prior reactions that precipitated DIC as mineral carbonates. Furthermore, DIC speciates primarily as bicarbonate or carbonate ions under high pH conditions such as those present in NSHQ14 [8]. Although methanogens may convert bicarbonate to CO₂ for use in methanogenesis or biomass generation, this reaction (HCO₃[−] + H⁺ → CO₂ + H₂O) consumes protons, which are limiting at high pH. Formatotrophic methanogenesis may help circumvent carbon acquisition problems because the oxidation of formate yields intracellular CO₂ which could be subsequently reduced to CH₄ (Fig. 2B) or biomass via the Wood-Ljungdahl pathway (Table S2). In addition, a

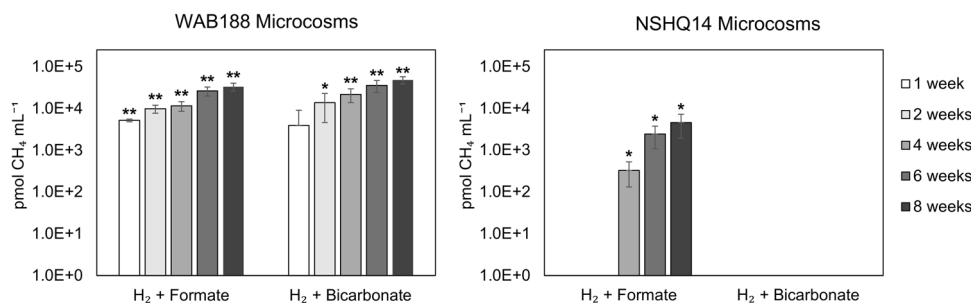


Fig. 4 Potential rates of biological methanogenesis from formate and bicarbonate by planktonic microbial communities in well water samples collected from the Samail Ophiolite in 2020. Potential rates of biological substrate transformation were determined via microcosm assays using well waters collected from the Samail Ophiolite in 2020. Results are plotted on a logarithmic scale. The average rates of methane generation observed in four replicate

abiological controls were subtracted from values in four replicate biological assays (Avg) and their combined standard deviations (SD) are presented at five timepoints over an 8-week incubation. *P*-values were determined between biological assays and abiological controls at each timepoint via Student's *t*-test assuming unequal variance for each condition (**p* < 0.05, ***p* < 0.01).

putative acetogen lacking all known hydrogenases and apparently using CO to drive acetogenesis was detected in the CO₂-limited, serpentinization-impacted groundwater of The Cedars, California [87]. This suggests that replacement of hydrogen-based reductant by single-carbon reduced substrates (i.e., formate or CO) may be a common strategy to circumvent CO₂ limitation among organisms encoding the Wood–Ljungdahl pathway of carbon fixation in serpentinizing environments.

The source of formate in waters of the Samail Ophiolite has not yet been resolved, however, consumption of CO₂ and H₂, coinciding with generation of nearly 100 μM formate, has been observed during low-temperature abiotic reactions between water and rocks from the Samail Ophiolite [38]. A positive correlation (Pearson *R* = 0.72, *p* < 0.05) existed between formate concentrations and pH in subsurface waters from the Samail Ophiolite sampled in 2015 [19] (formate was not measured in 2017), which is potentially consistent with the hypothesis that formate can be generated abiotically as a result of equilibration of CO₂ with H₂ produced during serpentinization reactions [8, 9, 88, 89]. In addition, formate formed via past reactions could potentially be stored in the rock (e.g., in fluid inclusions) [90]. Regardless of the source of formate in the system, it is possible that methanogens and other formotrophic populations actively maintain formate at a low (~1 μM) steady state concentration (Table 1).

To begin to test these predictions, well waters were sampled again from WAB188 and NSHQ14 in February and March, 2020, for use in quantifying rates of ¹⁴C-CH₄ generation from ¹⁴C-radiolabeled substrates (¹⁴C-HCOO⁻ and ¹⁴C-HCO₃⁻). Temperature and pH values were similar between waters sampled from each well in 2017 and 2020 (Table 1). Microcosms containing waters from WAB188 that were amended with either H₂ + bicarbonate (including 5 μCi ¹⁴C-HCO₃⁻) or with H₂ + formate (including 5 μCi

¹⁴C-HCOO⁻) generated significantly more CH₄ than killed controls, and CH₄ generation attributable to biological processes increased over time (Fig. 4, Supplementary Table S5). No significant differences were observed between the amount of biologically-generated CH₄ in WAB188 microcosms amended with H₂ + bicarbonate versus H₂ + formate at each time point. This is potentially consistent with the hypothesis that *Methanobacterium* Type I, which was detected in high abundance in well WAB188 among 2017 metagenomes, functions as a canonical Class I methanogen (similar to other members of the Methanobacteria) which can commonly use H₂ + CO₂ or formate as methanogenic substrates [70].

In contrast, significantly higher quantities of CH₄ were generated in NSHQ14 microcosms amended with H₂ + formate (including 5 μCi ¹⁴C-HCOO⁻) than in killed controls, and biologically-generated CH₄ increased over time (Fig. 4, Supplementary Table S5). However, no CH₄ generation attributable to biology was observed in NSHQ14 microcosms amended with H₂ + bicarbonate (including 5 μCi ¹⁴C-HCO₃⁻). This potentially supports the genomic prediction that *Methanobacterium* Type II, which was detected in high abundance in NSHQ14 among 2017 metagenomes, can couple H₂ with formate but not DIC to drive methanogenesis via a pathway that is unique among characterized Methanobacteria. Further, the rates of methanogenesis were lower in NSHQ14 microcosms as compared with WAB188 microcosms, which is potentially consistent with the previous finding that rates of utilization of select single-carbon compounds were generally lower in hyperalkaline waters than alkaline waters, possibly due to extreme conditions imposed by high pH waters [18].

The short branch length separating the Type I and Type II MAGs and differences in their methanogenesis pathways due to gene loss/acquisition suggests that the Type II cells were derived from the Type I lineage. This prompted the

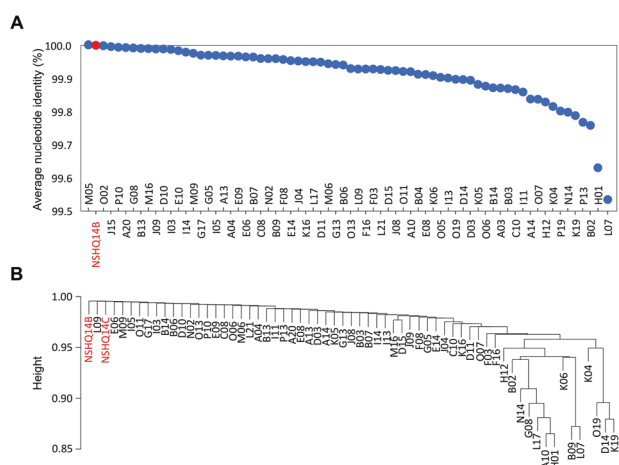


Fig. 5 Average nucleotide identities (ANIs) between *Methanobacterium* single amplified genomes (SAGs) and Type II meta-genome assembled genomes (MAGs). In **A**, ANIs relative to the NSHQ14C MAG are presented with blue circles representing individual SAGs and the red circle representing the NSHQ14B MAG. In **B**, pairwise ANIs among the SAGs and the NSHQ14C MAG were used to generate a dendrogram based on hierarchical clustering, with branch lengths (“Height”) representing Euclidean distances between ANIs calculated among the genomes. Red text represents the Type II MAGs and letters followed by two-digit numbers represent individual SAGs.

generation of single amplified genomes (SAGs) to further evaluate whether this phenomenon was prevalent among individual cells and to uncover evidence for continued strain-level diversification in this lineage. A total of 71 SAGs were assembled from NSHQ14C, 69 of which were affiliated with *Methanobacterium* (Supplementary Table S6). One of the other SAGs was affiliated with the methanotrophic bacterial genus, *Methyloccoccus*, and the other was too incomplete to classify taxonomically. The low genome recovery from SAGs (71 out of 317 single cells) is possibly attributable to the low efficiency of lysing cells that are adapted to hyperalkaline waters using potassium hydroxide.

Among the 69 *Methanobacterium* affiliated SAGs, the pairwise average nucleotide identities (ANIs) indicated an average difference of only ~1 bp per kbp between them and the NSHQ14C Type II MAG (average ANI of 99.90%) (Fig. 5A) and average amino acid identities between the SAGs and the NSHQ14C Type II MAG ranged from 95 to 100% (Supplementary Fig. S6). For comparison, the Type I MAG shared 91.11% and 90.59% ANI with the Type II MAGs from NSHQ14B and NSHQ14C, respectively. Proteins involved in cellular metabolism (e.g., HdrABC, FdhB, MvhD, and the large subunits of both Eha and Mrp–Mbh [NiFe]-hydrogenases) were all found to exhibit identical nucleotide sequences among the Type II SAGs and MAGs. Similarly, FdhA exhibited identical nucleotide sequences among all Type II SAGs and MAGs, except for a single SAG (B06) that was found to exhibit a single

non-synonymous nucleotide polymorphism resulting in a change from an aspartate to a glutamate. Both aspartate and glutamate have negatively charged side chains, indicating that this change likely conserved protein function, which is consistent with the presumed critical importance of FdhA in the metabolism of Type II cells (Fig. 3B). Consistent with the Type II MAGs, homologs of Mvh and Frh [NiFe]-hydrogenases were not found in any of the *Methanobacterium* affiliated SAGs. Together, these observations indicate that the core metabolism inferred for the Type II cells is likely a prevalent feature associated with this population.

In contrast, a number of orthologous proteins were identified in SAGs that differed markedly from those encoded by the NSHQ14C Type II MAG. Notably, among the ten orthologs that had the highest representation in SAGs and that differed from the NSHQ14C MAG, five were transposases (Table 2). This finding is in line with the previous detection of abundant transposases in a biofilm metagenome recovered from the Lost City serpentinizing system [91]. Mantel tests of matrices describing the ANI among SAGs (Supplementary Table S7) and the amino acid dissimilarity of the ten orthologs (Table 2) were used to identify orthologs that likely contributed to apparent diversification in the SAGs (Fig. 5). Among the ten orthologs considered, the dissimilarities in eight proteins were not significantly correlated with ANI. However, the dissimilarity of a protein belonging to the Peptidase C39 family and an ISNCY (“insertion sequence not classified yet”) transposase were significantly and positively correlated with the dissimilarity in ANI, suggesting that they may contribute to the apparent recent diversification of these cells.

The genes flanking the ten divergent orthologs were examined to identify evidence of recent transposition and/or horizontal gene transfer events in the form of different genomic locations in the SAGs. The unclassified transposase with the greatest number of SAG homolog variants that, based on the results of a Mantel test (Table 2), is not suggested to have significantly contributed to the recent vertical diversification of SAGs, was flanked by a randomized assortment of genes (data not shown). This is consistent with its recent and likely random integration into the genomes of Type II cells. In contrast, genes flanking the ISNCY family transposase that is suggested to have significantly contributed to recent diversification of these cells were largely syntenic among closely related orthologs but not among those encoding more divergent orthologs (Fig. 6). In addition, variability of gene order and gene content among the SAGs and MAGs suggested rearrangement of genes flanking this transposase. These observations suggest that multiple transposition events likely led to this variable genomic arrangement, thereby revealing a

Table 2 The ten most abundant orthologous proteins identified in single amplified genomes (SAGs) that differ from those in the NSHQ14C Type II metagenome assembled genome (MAG).

Variant protein ortholog annotation (SAGs vs. NSHQ14C MAG)	# of SAGs encoding orthologs with <100% identity to the NSHQ14C MAG	Amino acid identities (SAGs vs. NSHQ14C MAG)	Closest taxonomic affiliation of the most divergent ortholog	Mantel correlation (R) and significance (p) among protein ortholog amino acid dissimilarities and whole genome ANI dissimilarities
Transposase	21	96.8–98.4%	<i>Methanosarcina mazei</i>	$R = -0.01, p = 0.44$
GDP-mannose 4,6 dehydratase	17	73.9–100%	<i>Methanobacterium congolense</i>	$R = 0.07, p = 0.07$
DDE-type integrase/transposase/recombinase	17	97.2–100%	<i>Methanobacterium formicum</i>	$R = 0.09, p = 0.21$
Peptidase C39	16	76.5–100%	<i>Methanothermus fervidus</i>	$R = 0.19, p = 0.02$
Transposase (IS5 family)	16	96.6–100%	<i>Methanoculleus taiwanensis</i>	$R = 0.04, p = 0.23$
ISNCY family transposase	15	91.6–100%	<i>Methanobacterium subterraneum</i>	$R = 0.37, p = 0.01$
Histidine kinase	13	99.7–100%	<i>Methanobacterium formicum</i>	$R = 0.22, p = 0.07$
Transposase (IS630 family)	12	97.4–100%	<i>Methanobacterium subterraneum</i>	$R = -0.08, p = 0.84$
MFS transporter	12	99.8–100%	<i>Methanobacterium paludis</i>	$R = 0.01, p = 0.43$
Exosome complex protein Rrp4	7	99.6–100%	<i>Methanobacterium congolense</i>	$R = 0.01, p = 0.50$

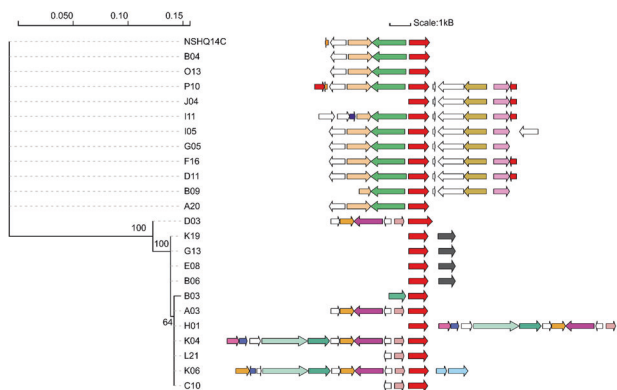


Fig. 6 ISNCY transposase protein phylogeny and genes colocalized with this transposase in the NSHQ14C metagenome assembled genome (MAG) and single cell genomes (SAGs). Branch length is relative to the scale provided at the top of the figure indicating average substitutions per site. Bootstrap values are displayed at each node (out of 100 replicates). Contigs that encoded this transposase in each SAG or MAG are depicted to the right of the terminals. Red arrows represent genes encoding transposase orthologs, white arrows represent genes encoding hypothetical proteins, black arrows represent genes encoding CRISPR-associated (Cas) proteins, and the remaining colored arrows represent genes encoding other functional proteins. Gene lengths are relative to the scale provided at the top of the figure.

mechanism that could have contributed to the recent diversification of Type II cells.

Intriguingly, SAGs K19, G13, E08, and B06 all encoded closely related ISNCY transposases upstream of a CRISPR-associated Cas4 protein (Fig. 6). This Cas4 protein was not encoded by the Type II NSHQ14B or NSHQ14C MAGs, with the most similar protein in the MAGs only exhibiting <30% identity. Moreover, this protein was only encoded by these four SAGs and one other (D03) where it was located on a contig that encoded multiple putative transposases (data not shown). Surprisingly, however, the contigs encoding Cas4 and the ISNCY transposase did not code for other Cas proteins or other genetic elements (i.e., CRISPRs and spacer sequences) that act in tandem with Cas4 [92] to confer adaptive immunity against invading viruses and plasmids in Bacteria and Archaea [93]. However, Cas4 is not always directly colocalized with other elements of CRISPR arrays [94] and, importantly, the ends of contigs represent portions of the genome where reads do not sufficiently overlap to generate contiguous sequences. Therefore, it is possible that transposases (especially ISNCY type) catalyzed the integration of functional CRISPR-Cas systems that may confer a subset of the population with adaptive immunity from foreign genetic material.

The findings herein are potentially in line with previous studies of organisms that are closely related to one another, wherein microbial genomes have been conceptualized as

containing two distinct components, core and variable, collectively termed the pan-genome [95, 96]. The core genome is composed of genes that are common to all strains, whereas the variable genome consists of genes that differ between strains, for example resulting from gain via horizontal gene transfer or loss. For example, seven *Sulfolobus islandicus* genomes recovered from globally distributed locations that diverged recently in evolutionary time (~910,000 years ago) shared a core genome including housekeeping genes, but also encoded variable genomic regions containing small inversions and rearrangements, many of which were also associated with CRISPR-Cas genes [95]. Like the *Methanobacterium* populations from 85 m depth in NSHQ14, other subsurface communities include population level genomic diversity that was proposed to have been generated via the activity of mobile elements [97–99].

Similar to the ISNCY transposase, genes flanking a protein belonging to the C39 peptidase family that is suggested to have contributed to recent diversification of Type II *Methanobacterium* cells (Supplementary Table 2) were largely syntenic among closely related proteins but not among those encoding more divergent orthologs, and rearrangement of genes co-localized with this peptidase were also observed (Supplementary Fig. S7). These rearrangements may have been catalyzed by a transposase that was detected downstream of the peptidase in at least one of these SAGs (I11). The diversification of peptidase C39 and its potential integration into cellular genomes by transposase activity may have implications for cellular fitness as these proteins function in bacteriocin (antimicrobial peptide) processing [100]. Thus, it is possible that Type II *Methanobacterium* cells employ bacteriocidin-type antimicrobials to increase their competitive advantage for limiting nutrients (e.g., formate) over other co-inhabiting species, potentially helping to explain their prevalence in hyperalkaline waters.

Conclusions

Reconstruction of two *Methanobacterium* genomes from metagenomic sequences obtained from subsurface waters exhibiting contrasting geochemical characteristics of the Samail Ophiolite, Sultanate of Oman, provided an opportunity to conduct comparative genomic and evolutionary investigations into adaptations that enable putative hydrogenotrophic methanogens to overcome DIC limitation associated with hyperalkaline conditions. Metabolic reconstruction of a Type I *Methanobacterium* MAG that was abundant in circumneutral, DIC replete waters revealed canonical hydrogenotrophic methanogenesis pathways. However, Type II *Methanobacterium* MAGs from

extensively serpentinized, DIC limited waters lacked homologs of key [NiFe]-hydrogenases, Mvh and Frh, that supply reductant from H₂ oxidation for methanogenesis. Metabolic reconstructions indicated that these functionalities were replaced by formate dehydrogenases that supply reductant from formate oxidation and that yield intracellular CO₂ to allow for methanogenesis to proceed under otherwise DIC limited conditions. The genomic prediction that *Methanobacterium* Type II can couple H₂ with formate but not DIC to drive methanogenesis via a pathway that is unique among characterized Methanobacteria was supported by microcosm-based radiotracer experiments revealing significant biological methane production from H₂ + formate but not H₂ + bicarbonate. The combination of phylogenetic and gene distribution data relative to other Methanobacteria indicated that the Type II lineage was derived from the Type I lineage, suggesting that the replacement of H₂-based reductant by formate-based reductant is a derived trait. In turn, this suggested that the directionality of the diversification of *Methanobacterium* inhabiting the Samail Ophiolite was from circumneutral into hyperalkaline environments, the latter of which is likely to be limiting in DIC, but where formate could be supplied by serpentinization reactions. Thus, diversification may have taken place in a system where serpentinization was generating large amounts of formate which could act as a selective pressure to evolve the suite of physiological traits allowing for its use in key methanogenesis reactions.

The recent diversification of *Methanobacterium* into hyperalkaline environments appeared to have been facilitated, in part, by lateral gene transfer of a [NiFe]-hydrogenase complex from a bacterium, perhaps affiliated with an ancestor of Firmicutes. Single cell genomes of Type II *Methanobacterium* revealed evidence of gene rearrangement, possibly through transposition. Collectively, these results point to the importance of gene loss, gain, and transposition in the adaptation of methanogens to hyperalkaline conditions and provide strong, albeit indirect, evidence that cells are active and adapted to the polyextremophilic subsurface environment of the serpentinizing Samail Ophiolite.

Acknowledgements This work was supported by a grant from the NASA Astrobiology Institute (NNA15BB02A) to JRS, AST, and ESB, and by NSF grants OCE-1335810 and OIA-1826734 to RS. We thank staff of the Single Cell Genomics Center at the Bigelow Laboratory for Ocean Sciences for generating the single cell genomics data. The authors are grateful to Juerg Matter for help with equipment acquisition, permitting, and sample export and to the Ministry of Regional Municipalities and Water Resources in the Sultanate of Oman for allowing sampling and export of well waters.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

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

References

- Boyd ES, Amenabar MJ, Poudel S, Templeton AS. Bioenergetic constraints on the origin of autotrophic metabolism. *Philos Trans R Soc A*. 2020;378:1471–2962.
- Boyd ES, Schut GJ, Adams MWW, Peters JW. Hydrogen metabolism and the evolution of biological respiration. *Microbe*. 2014;9:361–7.
- Hoehler TM. Biogeochemistry of dihydrogen (H₂). In: Sigel H, and Sigel R (eds.). *Metal ions in biological systems*. Vol 43. (Taylor & Francis Group, Boca Raton, FL, 2005) pp 9–48.
- Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. 2016;1:1–8.
- McCollom TM, Klein F, Robbins M, Moskowitz B, Berquó TS, Jöns N, et al. Temperature trends for reaction rates, hydrogen generation, and partitioning of iron during experimental serpentinization of olivine. *Geochim Cosmochim Acta*. 2016;181:175–200.
- Schulte M, Blake D, Hoehler T, McCollom T. Serpentinization and its implications for life on the early Earth and Mars. *Astrobiology*. 2006;6:364–76.
- Russell M, Hall A, Martin W. Serpentinization as a source of energy at the origin of life. *Geobiology*. 2010;8:355–71.
- Seewald JS, Zolotov MY, McCollom T. Experimental investigation of single carbon compounds under hydrothermal conditions. *Geochim Cosmochim Acta*. 2006;70:446–60.
- McCollom TM, Seewald JS. Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. *Chem Rev*. 2007;107:382–401.
- Twing KI, Brazelton WJ, Kubo MDY, Hyer AJ, Cardace D, Hoehler TM, et al. Serpentinization-influenced groundwater harbors extremely low diversity microbial communities adapted to high pH. *Front Microbiol*. 2017;8:308.
- Brazelton WJ, Nelson B, Schrenk MO. Metagenomic evidence for H₂ oxidation and H₂ production by serpentinite-hosted subsurface microbial communities. *Front Microbiol*. 2012;2:268.
- Morrill PL, Brazelton WJ, Kohl L, Rietze A, Miles SM, Kavanagh H, et al. Investigations of potential microbial methanogenic and carbon monoxide utilization pathways in ultra-basic reducing springs associated with present-day continental serpentinization: the Tablelands, NL, CAN. *Front Microbiol*. 2014;5:613.
- Crespo-Medina M, Twing KI, Sánchez-Murillo R, Brazelton WJ, McCollom TM, Schrenk MO. Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. *Front Microbiol*. 2017;8:916.
- Woychese KM, Meyer-Dombard DR, Cardace D, Argayosa AM, Arcilla CA. Out of the dark: transitional subsurface-to-surface microbial diversity in a terrestrial serpentinizing seep (Manleluag, Pangasinan, the Philippines). *Front Microbiol*. 2015;6:44.
- Neubeck A, Sun L, Müller B, Ivarsson M, Hosgörmez H, Özcan D, et al. Microbial community structure in a serpentine-hosted abiotic gas seepage at the Chimaera Ophiolite, Turkey. *Appl Environ Microbiol*. 2017;83:e03430–16.
- Lang SQ, Früh-Green G, Bernasconi SM, Brazelton WJ, Schrenk MO, McGonigle JM. Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City hydrothermal field. *Sci Rep*. 2018;8:1–10.
- Brazelton WJ, Morrill PL, Szponar N, Schrenk MO. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. *Appl Environ Microbiol*. 2013;79:3906–16.
- Fones EM, Colman DR, Kraus EA, Nothaft DB, Poudel S, Rempfert KR, et al. Physiological adaptations to serpentinization in the Samail Ophiolite, Oman. *ISME J*. 2019;13:1750–62.
- Rempfert KR, Miller HM, Bompard N, Nothaft D, Matter JM, Kelemen P, et al. Geological and geochemical controls on subsurface microbial life in the Samail Ophiolite, Oman. *Front Microbiol*. 2017;8:56.
- Kelemen PB, Matter J, Streit EE, Rudge JF, Curry WB, Blusztajn J. Rates and mechanisms of mineral carbonation in peridotite: natural processes and recipes for enhanced, in situ CO₂ capture and storage. *Annu Rev Earth Planet Sci*. 2011;39:545–76.
- Canovas PA, Hoehler T, Shock EL. Geochemical bioenergetics during low-temperature serpentinization: an example from the Samail ophiolite, Sultanate of Oman. *J Geophys Res*. 2017;122:1821–47.
- Suzuki S, Ishii S, Wu A, Cheung A, Tenney A, Wanger G, et al. Microbial diversity in The Cedars, an ultrabasic, ultrareducing, and low salinity serpentinizing ecosystem. *Proc Natl Acad Sci USA*. 2013;110:15336–41.
- Brazelton WJ, Thornton CN, Hyer A, Twing KI, Longino AA, Lang SQ, et al. Metagenomic identification of active methanogens and methanotrophs in serpentinite springs of the Voltri Massif, Italy. *Peer J*. 2017;5:e2945.
- Morrill PL, Kuenen JG, Johnson OJ, Suzuki S, Rietze A, Sessions AL, et al. Geochemistry and geobiology of a present-day serpentinization site in California: The Cedars. *Geochim Cosmochim Acta*. 2013;109:222–40.
- Miller HM, Matter JM, Kelemen P, Ellison ET, Conrad ME, Fierer N, et al. Modern water/rock reactions in Oman hyperalkaline peridotite aquifers and implications for microbial habitability. *Geochim Cosmochim Acta*. 2016;179:217–41.
- Russell MJ, Martin W. The rocky roots of the acetyl-CoA pathway. *Trends Biochem Sci*. 2004;29:358–63.
- Martin WF, Weiss MC, Neukirchen S, Nelson-Sathi S, Sousa FL. Physiology, phylogeny, and LUCA. *Microbiol. Cell*. 2016;3:582–7.
- Ueno Y, Yamada K, Yoshida N, Maruyama S, Isozaki Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature*. 2006;440:516–9.
- Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nat Geosci*. 2017;10:629–36.
- Etiopo G, Vadillo I, Whitaric MJ, Marques JM, Carreira PM, Tiago I, et al. Abiotic methane seepage in the Ronda peridotite massif, southern Spain. *Appl Geochem*. 2016;66:101–13.
- Proskurowski G, Lilley MD, Seewald JS, Früh-Green G, Olson EJ, Lupton JE, et al. Abiogenic hydrocarbon production at Lost City hydrothermal field. *Science*. 2008;319:604–7.

32. Etiope G. Methane origin in the Samail ophiolite: Comment on “Modern water/rock reactions in Oman hyperalkaline peridotite aquifers and implications for microbial habitability”. *Geochim Cosmochim Acta*. 2017;197:467–70.
33. Miller HM, Matter JM, Kelemen P, Ellison ET, Conrad ME, Fierer N, et al. Reply to “Methane origin in the Samail ophiolite: Comment on ‘Modern water/rock reactions in Oman hyperalkaline peridotite aquifers and implications for microbial habitability’”. *Geochim Cosmochim Acta*. 2017;197:471–3.
34. Miller HM, Chaudhry N, Conrad ME, Markus B, Kopf SH, Templeton AS. Large carbon isotope variability during methanogenesis under alkaline conditions. *Geochim Cosmochim Acta*. 2018;237:18–31.
35. Bradley AS, Hayes JM, Summons RE. Extraordinary ¹³C enrichment of diether lipids at the Lost City Hydrothermal Field indicates a carbon-limited ecosystem. *Geochim Cosmochim Acta*. 2009;73:102–18.
36. Zwicker J, Birgel D, Bach W, Richo S, Smrzka D, Grasemann B, et al. Evidence for archaeal methanogenesis within veins at the onshore serpentinite-hosted Chimaera seeps, Turkey. *Chem Geol*. 2018;483:567–80.
37. Kraus EA, Stamps BW, Rempfert KR, Nothaft DB, Boyd ES, Matter JM, et al. Biological methane cycling in serpentinization-impacted fluids of the Samail ophiolite of Oman. *AGU Fall Meeting Abstracts*. 2018; (abstract #V13E-0139).
38. Miller HM, Mayhew LE, Ellison ET, Kelemen P, Kubo M, Templeton AS. Low temperature hydrogen production during experimental hydration of partially-serpentinized dunite. *Geochim Cosmochim Acta*. 2017;209:161–83.
39. Neal C, Stanger G. Hydrogen generation from mantle source rocks in Oman. *Earth Planet Sci Lett*. 1983;66:315–20.
40. Streit E, Kelemen P, Eiler J. Coexisting serpentine and quartz from carbonate-bearing serpentinized peridotite in the Samail Ophiolite, Oman. *Contrib Miner Petr*. 2012;164:821–37.
41. Chavagnac V, Monnin C, Ceuleneer G, Boulart C, Hoareau G. Characterization of hyperalkaline fluids produced by low-temperature serpentinization of mantle peridotites in the Oman and Ligurian ophiolites. *Geochem Geophys*. 2013;14:2496–522.
42. Mervine EM, Humphris SE, Sims KWW, Kelemen PB, Jenkins WJ. Carbonation rates of peridotite in the Samail Ophiolite, Sultanate of Oman, constrained through ¹⁴C dating and stable isotopes. *Geochim Cosmochim Acta*. 2014;126:371–97.
43. Kang DWD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. 2015;3:e1165.
44. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun*. 2017;8:84.
45. Colman DR, Lindsay MR, Boyd ES. Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat Commun*. 2019;10:1–13.
46. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *J Bioinform*. 2012;28:1033–4.
47. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of highquality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
48. Nguyen LT, Schmidt HA, von Haesler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
49. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haesler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587.
50. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119.
51. Seemann T. Prokka: rapid prokaryotic genome annotation. *J Bioinform*. 2014;30:2068–9.
52. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
53. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35:W182–5.
54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
55. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
56. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J*. 2016;10:761–77.
57. Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, et al. [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *BBA-Mol Cell Res*. 2015;1853:1350–69.
58. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2016;45: D200–3.
59. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14:e1005944.
60. R Core Team. R: a language and environment for statistical computing. Version 3.0.1. R Foundation for Statistical Computing. 2013.
61. Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a CLASsification of Mobile genetic Elements, update 2010. *Nucleic Acids Res*. 2010;38:D57–D61.
62. Lefort V, Longueville J-E, Gascuel O. SMS: smart model selection in PhyML. *Mol Biol Evol*. 2017;34:2422–4.
63. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
64. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS One*. 2010;5:e11147.
65. Harrison KJ, Crécy-Lagard V, Zallot R. Gene Graphics: a genomic neighborhood data visualization web application. *J Bioinform*. 2018;34:1406–8.
66. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, and O’Hara RB. vegan: community ecology package. R Foundation for Statistical Computing. 2015.
67. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;25:725–31.
68. Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A, Morrill PL, et al. Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars. *ISME J*. 2017;11:2584–98.
69. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology. *ISME J*. 2014;8:1553–65.

70. Thauer RK, Kaster AK, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol*. 2008;6:579–91.
71. Hendrickson EL, Leigh JA. Roles of coenzyme F₄₂₀-reducing hydrogenases and hydrogen-and F₄₂₀-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F₄₂₀ and production of hydrogen during methanogenesis. *J Bacteriol*. 2008;190:4818–21.
72. Goldman AD, Leigh JA, Samudrala R. Comprehensive computational analysis of Hmd enzymes and paralogs in methanogenic Archaea. *BMC Evol Biol*. 2009;9:199.
73. Tersteegen A, Hedderich R. Methanobacterium thermoautotrophicum encodes two multisubunit membrane-bound [NiFe] hydrogenases: transcription of the operons and sequence analysis of the deduced proteins. *Eur J Biochem*. 1999;264:930–43.
74. Lie TJ, Costa KC, Lupa B, Korpole S, Whitman WB, Leigh JA. Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. *Proc Natl Acad Sci USA*. 2012;109:15473–8.
75. Thauer RK. The Wolfe cycle comes full circle. *Proc Natl Acad Sci USA*. 2012;109:15084–5.
76. Costa KC, Wong PM, Wang T, Lie TJ, Dodsworth JA, Swanson I, et al. Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. *Proc Natl Acad Sci USA*. 2010;107:11050–5.
77. Greening C, Ahmed FA, Mohamed AE, Lee BM, Pandey G, Warden AC, et al. Physiology, biochemistry, and applications of F₄₂₀-and Fo-dependent redox reactions. *Microbiol Mol Biol Rev*. 2016;80:451–93.
78. Yan Z, Ferry JG. Electron bifurcation and confurcation in methanogenesis and reverse methanogenesis. *Front Microbiol*. 2018;9:1322.
79. Costa KC, Lie TJ, Xia Q, Leigh JA. VhuD facilitates electron flow from H₂ or formate to heterodisulfide reductase in *Methanococcus maripaludis*. *J Bacteriol*. 2013;195:5160–5.
80. Schauer NL, Ferry JG. Properties of formate dehydrogenase in *Methanobacterium formicicum*. *J Bacteriol*. 1982;150:1–7.
81. Schauer NL, Ferry JG, Honek JF, Orme-Johnson WH, Walsh C. Mechanistic studies of the coenzyme F₄₂₀-reducing formate dehydrogenase from *Methanobacterium formicicum*. *Biochemistry*. 1986;25:7163–8.
82. Mills DJ, Vitt S, Strauss M, Shima S, Vonck J. De novo modeling of the F₄₂₀-reducing [NiFe]-hydrogenase from a methanogenic archaeon by cryo-electron microscopy. *Elife*. 2013;2:e00218.
83. Schut GJ, Boyd ES, Peters JW, Adams MWW. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev*. 2013;37:182–203.
84. Hamamoto T, Hashimoto M, Hino M, Kitada M, Seto Y, Kudo T, et al. Characterization of a gene responsible for the Na⁺/H⁺ antiporter system of alkalophilic *Bacillus* species strain C125. *Mol Microbiol*. 1994;14:939–46.
85. Buckel W, Thauer RK. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na⁺ translocating ferredoxin oxidation. *BBA-Bioenerg*. 2013;1827:94–113.
86. Boone DR, Johnson RL, Liu Y. Diffusion of the interspecies electron carriers H₂ and formate in methanogenic ecosystems and its implications in the measurement of K_m for H₂ or formate uptake. *Appl Environ Microbiol*. 1989;55:1735–41.
87. Suzuki S, Neelson KH, Ishii S. Genomic and in-situ transcriptomic characterization of the candidate phylum NPL-UPL2 from highly alkaline highly reducing serpentinized groundwater. *Front Microbiol*. 2018;9:3141.
88. Lang SQ, Butterfield DA, Schulte M, Kelley DS, Lilley MD. Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim Cosmochim Acta*. 2010;74:941–52.
89. McCollom TM, Seewald JS. Experimental constraints on the hydrothermal reactivity of organic acids and acid anions: I. Formic acid and formate. *Geochim Cosmochim Acta*. 2003;67:3625–44.
90. Zeng Y, Liu J. Short-chain carboxylates in fluid inclusions in minerals. *Appl Geochem*. 2000;15:13–25.
91. Brazelton WJ, Baross JA. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J*. 2009;3:1420–4.
92. Zhang J, Kasciukovic T, White MF. The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One*. 2012;7:e47232.
93. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*. 2015;117:119–28.
94. Jansen R, van Embden JDA, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002;43:1565–75.
95. Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc Natl Acad Sci USA*. 2009;106:8605–10.
96. Tettelin H, Massignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome.”. *Proc Natl Acad Sci USA*. 2005;102:13950–5.
97. Labonté JM, Field EK, Lau M, Chivian D, Van Heerden E, Wommack KE, et al. Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Front Microbiol*. 2015;6:349.
98. Karnachuk OV, Frank YA, Lukina AP, Kadnikov VV, Beletsky AV, Mardanov AV, et al. Domestication of previously uncultivated *Candidatus Desulforudis audaxviator* from a deep aquifer in Siberia sheds light on its physiology and evolution. *ISME J*. 2019;13:1947–59.
99. Paul BG, Burstein D, Castelle CJ, Handa S, Arambula D, Czornyj E, et al. Retroelement-guided protein diversification abounds in vast lineages of bacteria and archaea. *Nat Microbiol*. 2017;2:17045.
100. Dirix G, Monsieus P, Dombrecht B, Daniels R, Marchal K, Vanderleyden J, et al. Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. *Peptides*. 2004;25:1425–40.